JOURNAL OF APPROXIMATION THEORY 63, 56–64 (1990)

Zeros of Chebyshev Polynomials in Markov Systems

PETER BORWEIN*

Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, N.S., Canada B3H 3J5

Communicated by Paul Nevai

Received January 3, 1989

We prove, under certain additional assumptions, that a Markov system has dense span if and only if the zeros of the associated Chebyshev polynomials are dense.

1. INTRODUCTION

Our main purpose is to prove that, under certain reasonable restrictions, a Markov system is dense (that is, it has dense span) if and only if the zeros of the associated Chebyshev polynomials are dense. This is the content of Theorem 1. This relationship has been conjectured by various people including von Golitschek, Kroó, Saff, and the author [5]. Half of this, namely that density of the system implies density of the zeros, is established in [5] and is done in a more general setting than ours. For completeness and simplicity we also offer an easy proof of this direction of the conjecture. Theorem 2 relates the behavior of the zeros to the rate of approximation. The final section of the paper examines the Muntz case in a little additional detail.

The notations we need are the following: An infinite *Markov system* on an interval $[a, b]$ is a collection of continuous functions on $[a, b]$

$$\mathscr{M} := \{g_1 := 1, g_2, g_3, g_4, \ldots\} \tag{1.1}$$

with the property that if an element of the real linear span of the first n, i.e., an element of

$$H_n := \operatorname{Span}\{g_1, g_2, \ldots, g_n\}, \tag{1.2}$$

* Research supported by N.S.E.R.C. of Canada.

0021-9045/90 $3.00

vanishes at n points, then it vanishes identically. This latter condition is called the Haar condition. Note that 1 is always an element of our Markov system.

Standard examples on $[0, 1]$ are

$$\{1, x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \ldots\}, \tag{1.3}$$

where the λ_i are distinct positive numbers and

$$\left\{1, \frac{1}{x+\lambda_1}, \frac{1}{x+\lambda_2}, \frac{1}{x+\lambda_3}, \ldots\right\}. \tag{1.4}$$

The fact that $\{g_1, \ldots, g_n\}$ satisfies the Haar condition on $[a, b]$ guarantees that unique best approximations exist in the uniform norm to a continuous f on any closed subset $X \subset [a, b]$. Furthermore, the best approximation, $p \in H_n$, is characterized by the *alternation property* that there exist $n+1$ points $x_i \in X$ with $x_i < x_{i+1}$ and with

$$f(x_i) - p(x_i) = -[f(x_{i+1}) - p(x_{i+1})] = \pm \|f(x) - p(x)\|_X, \tag{1.5}$$

where $\|\cdot\|_X$ denotes the sup norm on X. In fact, existence of unique best approximations from H_n is equivalent to H_n satisfying the Haar condition. This standard theory may be found in [2] or [4].

The additional assumption we will at times place on our Markov system is

Assumption 1. We say that an infinite Markov system on $[a, b]$,

$$\mathcal{M} := \{g_1 := 1, g_2, g_3, g_4, \ldots\},$$

satisfies Assumption 1 if each g_i is differentiable on (a, b) and if $f \in H_n$ and f' has $n-1$ zeros on (a, b) then f is identically constant.

Both of the systems (1.3) and (1.4) satisfy Assumption 1. So does any Markov system, that contains 1, where the derivatives also form a Markov system. Note that the derivatives are not assumed to be continuous.

Assumption 1 is not as strong as it looks. In fact, any Markov system (with $g_1 := 1$) that is composed of C^∞ functions satisfies Assumption 1. (See [4, p. 378].)

The (*general*) *Chebyshev polynomial*, T_n, associated with the Markov system on $[a, b]$ is the linear form

$$T_n := c\left[g_n - \sum_{k=1}^{n-1} c_i g_i\right], \tag{1.6}$$

where the c_i are chosen so that $\sum_{k=1}^{n-1} c_i g_i$ is the best approximation to g_n from H_{n-1} and where c is chosen so that $\|T_n\|_{[a,b]} = 1$ and $T_n(b) > 0$. This uniquely defines T_n. The properties T_n shares with the usual Chebyshev polynomial (of degree $n-1$) are that T_n has exactly $n-1$ zeros on $[a, b]$ and that T_n oscillates between ± 1 exactly n times on $[a, b]$. Also the zeros of T_{n-1} interlace the zeros of T_n. The oscillation property is merely a result of being a best approximation. The interlacing of zeros follows on consideration of $T_n \pm T_{n-1}$ which would have at least n zeros if the zeros did not interlace. Finally, let

$$Z_{\mathscr{M}} := \{x \in [a, b] \mid T_n(x) = 0 \text{ for some } n\} \tag{1.7}$$

denote the set of zeros of the Chebyshev polynomials and let

$$M_n := \max_{1 \leqslant i \leqslant n} |x_i - x_{i-1}|, \tag{1.8}$$

where $x_1 < x_2 < \cdots < x_{n-1}$ are the zeros of T_n and where $x_0 := a$ and $x_n := b$. (This is the maximum length of a zero free interval of T_n.) From the interlacing of the zeros of T_n it is easy to deduce that

$$M_n \to 0 \qquad \text{iff} \quad \underline{\lim}\, M_n = 0. \tag{1.9}$$

We say that $Z_{\mathscr{M}}$ is dense in $[a, b]$ if $\underline{\lim}\, M_n = 0$ (this is the appropriate notion of denseness for a sequence of sets).

When we say that a Markov system is dense we mean that the closed linear span of $\mathscr{M}$ is dense in the supremum norm in the continuous functions on $[a, b]$. Our principal result which we prove in the next section now states that, subject to Assumption 1, $\mathscr{M}$ is dense if and only if $Z_{\mathscr{M}}$ is dense in $[a, b]$.

The theory of Markov systems may be accessed in [2, 4]. Papers that examine the relationship between best uniform approximations and locations of zeros or extrema of best approximations are [1, 5, 6].

2. Denseness

We proceed to show that denseness of zeros of the Chebyshev polynomials implies denseness of the Markov system. Our approach is to construct approximate step functions. In particular we will show, under the assumption that $Z_{\mathscr{M}}$ is dense, that given $[c, d] \subset [a, b]$ and $\varepsilon > 0$ there exists $S(x) \in \operatorname{Span}\{g_1, g_2, ...\}$ such that

$$\begin{aligned} |S(x)| &\leqslant \varepsilon && x \in [a, c] \\ |S(x) - 1| &\leqslant \varepsilon && x \in [d, b] \\ S(x)' &\geqslant 0 && x \in [c, d]. \end{aligned} \tag{2.1}$$

Given that such a class of S functions exists showing denseness of $\mathcal{M}$ is now easy and standard. The argument is roughly as follows. If $\mathcal{M}$ were not dense there would exist a non-trivial continuous linear functional, which in this case by the Riesz representation theorem is a Borel measure μ, vanishing on the closed span of $\mathcal{M}$. In particular

$$\int_a^b S(x)\,d\mu(x) = 0 \tag{2.2}$$

for all S functions of the preceding type. This, however, implies that μ vanishes identically.

Alternatively one may explicitly construct approximants as in Theorem 2. The construction of S functions is the content of the next lemma.

LEMMA 1. *Let $\mathcal{M}$ be any Markov system on $[a, b]$ that satisfies Assumption* 1. *Let $[c, d] \subset [a, b]$. Suppose $S_n \in H_n$ is a best approximant from H_n to f on $[a, c] \cup [d, b]$, where*

$$f(x) := \begin{cases} 0, & x \in [a, c] \\ 1, & x \in [d, b]. \end{cases}$$

Then

(A) *S_n is monotone on $[c, d]$.*

(B) *If $Z_{\mathcal{M}}$ is dense in $[c, d]$ (that is, if the mesh of the zeros in $[c, d]$ tends to zero) then* $\lim \|f - S_n\|_{[a,c] \cup [d,b]} = 0$.

(C) $\|S_n - f\|_{[a,c] \cup [d,b]} \leq 10M_n/(d-c)$.

Proof. Since S_n is a best approximant to f there exist $n+1$ points where the maximum error, ε_n, occurs with alternating sign. Suppose $m+1$ of these points $x_0, ..., x_m$ lie in $[a, c]$ and $n-m$ of these points $x_{m+1}, ..., x_n$ lie in $[d, b]$. Then S_n' has at least $m-1$ zeros in (a, c) (one at each alternation point in $[a, c]$ except possibly at the endpoints a and c). Likewise S_n' has at least $n-m-2$ zeros in (d, b). So S_n' has at least $n-3$ zeros in $(a, c) \cup (d, b)$. Note that this count excludes x_m and x_{m+1}. Thus S_n' has at most one more zero in (a, b) unless S_n' vanishes (which is only possible for $n=1$). Now suppose S_n' has a zero (with sign change) on (c, d). Then since there is at most one zero of S_n' in (c, d) it cannot be the case that both $x_m := c$ and $x_{m+1} := d$ are alternation points with both $S_n'(c) \neq 0$ and $S_n'(d) \neq 0$. (Otherwise $\operatorname{sign}(S_n(c) - f(c)) = \operatorname{sign}(S_n(d) - f(d))$ as a consideration of the two cases shows.) But if $x_m \neq c$ or $x_{m+1} \neq d$ or $S_n'(c) = 0$ or $S_n'(d) = 0$ we have accounted for all the zeros of S_n' by accounting for the one (possibly) additional zero (either S_n' vanishes at c or d or one of x_m or x_{m+1} is an interior alternation point where S_n' vanishes). Thus S_n' has no zeros in (c, d) and (A) is proved.

For part (B) we make the following observation. Let

$$\varepsilon_n := \|f - S_n\|_{[a,c] \cup [d,b]}.$$

Then

$$D_n := \varepsilon_n T_n - S_n$$

has at least $m-1$ zeros on $[a, c]$ and

$$D_n^* := D_n + 1 = 1 + \varepsilon_n T_n - S_n$$

has at least $n-m-2$ zeros on $[d, b]$. Thus D_n' has at least $n-3$ zeros on $[a, c] \cup [d, b]$. Suppose T_n, the nth Chebyshev polynomial on $[a, b]$, has at least 4 alternations on an interval $[\alpha, \beta] \subset (c, d)$ and suppose that

$$S_n(\beta) - S_n(\alpha) \leqslant \varepsilon_n.$$

Then, because of the oscillation of T_n on $[\alpha, \beta]$,

$$D_n + \frac{S_n(\beta) + S_n(\alpha)}{2} = \varepsilon_n T_n - \left[S_n - \frac{S_n(\beta) + S_n(\alpha)}{2} \right]$$

has at least 3 zeros on $[\alpha, \beta]$ and hence

$$D_n' = \left(D_n + \frac{S_n(\beta) + S_n(\alpha)}{2} \right)'$$

has at least 2 zeros on $[\alpha, \beta]$. This, however, gives D_n' a total of at least $n-1$ zeros, which is impossible. In particular

$$S_n(\beta) - S_n(\alpha) > \varepsilon_n$$

on any interval $[\alpha, \beta] \subset (c, d)$ where T_n has at least 4 alternations. However, for any fixed k for large n, T_n has at least k such intervals (since the largest gap in the zeros of T_n on $[c, d]$, M_n, tends to zero by the assumption on the density of the zeros). Thus with monotonicity

$$S_n(d) - S_n(c) \geqslant k\varepsilon_n \qquad \text{for large } n.$$

However, by construction

$$S_n(d) - S_n(c) \leqslant 1 + 2\varepsilon_n$$

and ε_n must tend to zero.

For part (C) observe that k, as above, may be chosen to be $(d-c)/5M_n$. Also note that $\varepsilon_n \leqslant \frac{1}{2}$. So comparison of the two inequalities above yields the result. ∎

THEOREM 1. (A) *Suppose that $\mathcal{M}$ is an infinite Markov system on $[a, b]$ and suppose that $Z_{\mathcal{M}}$ is not dense. Then $\mathcal{M}$ is not dense.*

(B) *Suppose that $\mathcal{M}$ is an infinite Markov system on $[a, b]$ which satisfies Assumption 1, and suppose that $Z_{\mathcal{M}}$ is dense. Then $\mathcal{M}$ is dense.*

Proof. We offer the following simple proof of (A). Suppose $[c, d]$ contains no zero of T_n. Consider the piecewise linear function F defined as follows. Let $c < x_1 < x_2 < x_3 < x_4 < d$, and let

$$F(x) := \begin{cases} 0, & x = a, c, b, d \\ 2, & x = x_1, x_3 \\ -2, & x = x_2, x_4 \end{cases}$$

and be linear elsewhere. Suppose there exists $p \in H_n$ with

$$\|p(x) - F(x)\|_{[a,b]} < \tfrac{1}{2}. \tag{2.3}$$

Then

$$p(x) - T_n(x)$$

has $n-3$ zeros on $[a, c] \cup [d, b]$ because T_n has at least $n-1$ extrema on these intervals, but the four extrema of p on (c, d) guarantee three more zeros on this interval. Hence $p - T_n$ has n zeros and vanishes identically. This contradicts (2.3).

Part (B) is just a matter of coupling the argument that begins this section with part C of the lemma. See also Theorem 2.

Note that if $\mathcal{M}$ is a Markov system on $[a, b]$ then it is also a Markov system on any subset. Thus, in particular, under Assumption 1, $\mathcal{M}$ is dense in $C[\alpha, \beta]$ where $[\alpha, \beta]$ is any interval on which $Z_{\mathcal{M}}$ is dense.

THEOREM 2. *Let $\mathcal{M}$ be a Markov system on $[a, b]$ that satisfies, for all $m \in \mathbb{N}$, Assumption 1. Let $f \in C[a, b]$ and let p_n be the best approximation to f from H_n. Then*

$$\|p_n - f\|_{[a,b]} \leqslant D[1 + m^2 M_n]\, w_f\left(\frac{1}{m}\right),$$

where w_f is the modulus of continuity of f on $[a, b]$ and D depends only on a and b.

Proof. For simplicity suppose the interval in question is [0, 1]. The idea is to approximately interpolate f by a combination of S functions at the points $1/m, 2/m, \ldots, (m-1)/m$. Let ρ_i be defined by

$$\rho_i := \begin{cases} 0, & x \in \left[0, \dfrac{i}{m}\right] \\ 1, & x \in \left[\dfrac{i+1}{m}, 1\right] \end{cases}$$

and let $r_i \in H_n$ be defined (as in Lemma 1) to be the best approximation from H_n to ρ_i on $[0, i/m] \cup [(i+1)/m, 1]$. Let

$$U(x) := \sum_{i=0}^{m-1} \left(f\left(\frac{i+1}{m}\right) - f\left(\frac{i}{m}\right)\right) r_i(x) + f(0).$$

Then

$$U(x) \in H_n$$

and from Lemma 1 (C) with $(d-c) = 1/m$,

$$|U(x) - f(x)| \leq m^2 10 M_n w_f\left(\frac{1}{m}\right) + w_f\left(\frac{1}{m}\right). \quad \blacksquare$$

COROLLARY 1. *In the notation of the above theorem, if*

$$M_n = O\left(\frac{1}{n}\right)$$

then

$$\|p_n - f\|_{[a,b]} = O(w_f(n^{-1/2})).$$

3. THE MUNTZ CASE

We restrict our attention in this section to Markov systems

$$\mathscr{M} := \{1, x^{\lambda_1}, x^{\lambda_2}, \ldots\}, \tag{3.1}$$

where $1 < \lambda_1 < \lambda_2 < \cdots < \lambda_n \to \infty$. Then the classical theorem of Muntz says that $\mathscr{M}$ is dense on [0, 1] exactly when

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty. \tag{3.2}$$

In the case that $\mathcal{M}$ of (3.1) forms a non-dense Markov system we can say considerably more than just that $Z_{\mathcal{M}}$ is not dense. In fact we have the following.

THEOREM 3. *Suppose that $\mathcal{M}$, as in* (3.1), *is not dense. Then $\bar{Z}_{\mathcal{M}}$ is a countable set with second derived set $Z''_m = \{1\}$. In particular Z_m is nowhere dense.*

The proof of Theorem 3 rests on the following lemma which says that the Markov inequality for non-dense Muntz system is uniformly bounded except at 1.

LEMMA 2. *Suppose that $\mathcal{M}$, as in* (3.1), is *not dense. For each $\varepsilon > 0$ there exists η (independent of n) such that if $f \in H_n$ then*

$$\|f'(x)\|_{[0,1-\varepsilon]} \leqslant \eta \, \|f(x)\|_{[0,1]}.$$

Proof. In the case that $\mathcal{M}$ is not dense the closure of $\mathcal{M}$ is contained in the set of power series of the form

$$\sum_{i=0}^{\infty} a_i x^{\lambda_i}, \qquad \lambda_0 := 1,$$

where the above series converges on $[0, 1)$. This is a result of Clarkson and Erdős [3]. An examination of the proof shows that in this case the following inequality holds: If

$$f(x) := \sum_{i=0}^{n-1} a_i x^{\lambda_i} \in H_n$$

then there exists C_ρ depending only on $\rho > 0$ such that

$$|a_i| < C_\rho \, \|f\|_{[0,1]} (1+\rho)^i.$$

Then

$$f'(x) = \sum_{i=0}^{n-1} \lambda_i a_i x^{\lambda_i - 1}$$

and

$$|f'(x)| \leqslant C_\rho \, \|f\|_{[0,1]} \sum_{i=1}^{n-1} \lambda_i (1+\rho)^{\lambda_i} x^{\lambda_i - 1}$$

from which the lemma follows. ∎

Proof of Theorem 3. Let T_n be the nth Chebyshev polynomial with respect to $\mathcal{M}$. For any fixed ε, since T'_n is uniformly bounded on $[0, 1-\varepsilon]$, T_n can have at most some fixed number, say k_ε, of alternation points and hence zeros on $[0, 1-\varepsilon]$. Furthermore on $[0, 1-\varepsilon]$ consecutive zeros of T_n can be no closer than distance η_ε, where η_ε is also independent of n. This again follows from the uniformly bounded derivatives. Since the zeros of T_n interlace, the smallest cluster point in $Z_{\mathcal{M}}$ is the limit of the decreasing sequence of the smallest zeros of each T_n. The second smallest cluster point is the limit of the sequence of second smallest zeros, etc. There are at most k_ε such cluster points on $[0, 1-\varepsilon]$. ∎

A similar result holds for Markov systems of the form $\{1/(x+\alpha_i)\}_{i=1}^{\infty}$ where $\alpha_i \downarrow 0$ and it holds for essentially the same reasons as Theorem 3 holds. Thus in this case and the Muntz case either $Z_{\mathcal{M}}$ is dense or it is really quite thin and this in some way reflects the fact that either $\mathcal{M}$ is dense or it is really quite sparse. (Is this a general phenomenon? Does there exist a Markov system with $Z_{\mathcal{M}}$ neither dense nor nowhere dense?) Furthermore, these results relate intimately to the uniform boundedness of the corresponding Markov's inequality. The exact relationship between Markov's inequality and denseness also appears to warrant further study.

We finish with an observation on external points of best approximations.

THEOREM 4. *Suppose $\mathcal{M}$ is a non-dense Muntz system, and suppose f is not in the closure of the linear span of $\mathcal{M}$. Then the set of alternation points of the best approximations to f from each H_n form a set $\mathcal{A}_{\mathcal{M}}$ where $\mathcal{A}_{\mathcal{M}}$ is nowhere dense and $\mathcal{A}''_{\mathcal{M}} = \{1\}$.*

The proof is identical to that of Theorem 3; the functions

$$f - \rho_n,$$

where ρ_n is the best approximant to f from H_n, play the role of the T_n.

REFERENCES

1. P. B. BORWEIN, R. GROTHMANN, A. KROÓ, AND E. B. SAFF, The density of alternation points in rational approximations, *Proc. Amer. Math. Soc.* **105** (1989), 881–888.
2. E. W. CHENEY, "Introduction to Approximation Theory," McGraw-Hill, New York, 1966.
3. J. A. CLARKSON AND P. ERDŐS, Approximation by polynomials, *Duke Math. J.* **10** (1943), 5–11.
4. S. KARLIN AND W. J. STUDDEN, "Tchebycheff Systems with Applications in Analysis and Statistics," Wiley, New York, 1966.
5. A. KROÓ AND F. PEHERSTORFER, On the distribution of extremal points of general Chebyshev polynomials, in press.
6. G. G. LORENTZ, Distribution of alternation points in uniform polynomial approximation, *Proc. Amer. Math. Soc.* **92** (1984), 401–403.