JOURNAL OF APPROXIMATION THEORY **30**, 160–168 (1980)

Interpolation by Non-Negative Polynomials

JOHN BRIGGS AND LEE A. RUBEL*

Department of Computer Science and Department of Mathematics, University of Illinois, Urbana, Illinois 61801

Communicated by E. W. Cheney

Received April 9, 1979

DEDICATED TO THE MEMORY OF GÉZA FREUD

It is an elementary fact that given any continuous real function f on $[0, 1]$ and any $n + 1$ points on its graph, there exists a unique polynomial p_n of degree $\leqslant n$ that passes through those points. It is natural to ask whether, if f is non-negative, one may choose the $n + 1$ points on the graph of f so that the interpolating polynomial p_n is also non-negative. We prove here that this is indeed the case. (This problem was posed to us by Charles Chui.) Our proof is entirely elementary. The reader is urged to draw his own pictures to bring out the simple geometric nature of the proof.

For applications to approximation by splines, it would be useful, in the problem we do solve here, to make sure that the two endpoints of the interval are among the points of interpolation. But our method of proof is actually destructive of this goal—see for example the proof of Case D near the end of the paper. (See also the final remark of our paper.)

THEOREM. *Let f be a continuous and non-negative real function on the closed interval $[0, 1]$. For each $n = 0, 1, 2,\ldots,$ there exist $n + 1$ points $0 \leqslant x_0 < x_1 < \cdots < x_n \leqslant 1$ and a polynomial p_n of degree $\leqslant n$ that is also non-negative on $[0, 1]$, so that $p_n(x_k) = f(x_k)$ for $k = 0, 1,\ldots, n$.*

Remark. The corresponding theorem for periodic functions f and trigonometric polynomials p_n is also true. The proof is along the same lines, but is somewhat easier because of the absence of endpoint effects.

DEFINITION. Let f be continuous and real on $[0, 1]$ and let p be a polynomial such that $f - p$ has only finitely many zeros. Let $t \in [0, 1]$, and write $\Delta = f - p$. Then

* The research of the second author was partially supported by a grant from the National Science Foundation.

0021-9045/80/110160-09$02.00/0

(a) t is called an intersection value of (f, p) if $\Delta(t)=0$.

(b) t is called a crossover value of (f, p) if $\Delta(x)$ is strictly positive in $[t-h, t)$ and strictly negative in $(t, t+h]$, or vice versa, for some number $h>0$.

(c) t is called a kiss value of (f, p) if $\Delta(t)=0$ and $\Delta(x) \geqslant 0$ (or $\Delta(x) \leqslant 0$) for all x in $[t-h, t+h]$ for some number $h>0$.

(d) t is called an end intersection value of (f, p) if $\Delta(t)=0$ and either $t=0$ or $t=1$.

(e) t is called a good intersection value if it is either a crossover or an end intersection value.

Note. Once and for all, we exclude the case where $f(x)-p(x)$ has infinitely many intersection values, for in that case, we may trivially choose $p_n=p$ for all $n \geqslant \deg p$.

LEMMA 1 (Obvious). *The intersection values of (f, p) are of precisely three kinds*: (1) *crossover*, (2) *kiss*, *and* (3) *end*.

LEMMA 2. *If t is a crossover value of (f, p) and a positive number ε is given, then for every sufficiently small $\delta>0$, the following implication holds for all polynomials q. If $|p(x)-q(x)|<\delta$ for all $x \in [0,1]$, then (f, q) has at least one crossover value t' in $[t-\varepsilon, t+\varepsilon]$. We call this value t' a crossover value of (f, q) that corresponds to t. No confusion can arise if ε is small enough.*

Proof. Suppose, say, that $0<h<\varepsilon$ and that $p(t-h)<f(t-h)-2\delta$ and that $p(t+h)>f(t+h)+2\delta$. Then we would have $q(t-h)<f(t-h)-\delta$ and $q(t+h)>f(t+h)+\delta$, so that there must exist a crossover of (f, q) in $[t-h, t+h]$.

Proof of the Theorem

We shall suppose that f has only finitely many zeros, since we could otherwise take $p_n(x) \equiv 0$ for all n. We shall construct p_n inductively so that

(1) $p_n(x) \geqslant 0$ for all $x \in [0,1]$

(2) $\deg p_n(x) \leqslant n$

(3) there are at least $n+1$ good intersection values of (f, p_n) in $[0,1]$

(4) $p_n(x)$ is strictly positive for each $x \in (0,1)$.

It is easy to start the induction. Let

$$p_0(x)=\tfrac{1}{2}\left[\max_{y \in [0,1]} f(y)+\min_{y \in [0,1]} f(y)\right].$$

Further let

$$p_1(x) = (1-x)\, f(0) + xf(1),$$

provided that $f(0)$ and $f(1)$ are not both 0. Otherwise let

$$p_1(x) = \tfrac{1}{2} \max_{y \in [0,1]} f(y).$$

It is easy to see that p_0 and p_1 satisfy our conditions (1)–(4). Now suppose that p_n has been constructed and that $n \geqslant 1$.

If (f, p_n) has at least $n+2$ good intersection values, then we let $p_{n+1} = p_n$. Otherwise, there are exactly $n+1$ good intersection values x_i with $0 \leqslant x_0 < x_1 < \cdots < x_n \leqslant 1$. Since there are no extra cross-over values, then surely $f - p_n$ is alternately non-negative and non-positive in the successive open intervals between the x_j, where we adjoin the intervals $(0, x_0)$ and $(x_n, 1)$ if they are non-empty.

We now let

$$r_{n+1}(x) = (x - x_0)(x - x_1) \cdots (x - x_n),$$

and choose the polynomial $s_{n+1} = \pm\, r_{n+1}$ so that s_{n+1} and $f - p_n$ have the same sign away from their zeros, i.e.,

$$[f(x) - p_n(x)]\, s_{n+1}(x) \geqslant 0 \qquad \text{for all} \quad x \in [0, 1].$$

We further define

$$q_{n+1}(x, \varepsilon) = p_n(x) + \varepsilon s_{n+1}(x)$$

and we let

$$\alpha = \sup \{\varepsilon \geqslant 0\colon [f(x) - q_{n+1}(x, \varepsilon)]\, s_{n+1}(x) \geqslant 0 \text{ for all } x \in [0, 1]\}.$$

It is possible that $\alpha = 0$. The idea is that we are shifting p_n *towards* f until either a new good intersection value is created or is about to be created in the sense that for ε slightly larger than α, $(f(x), q_{n+1}(x, \varepsilon))$ will have at least one more good intersection value than (f, p_n) does. We let

$$q_{n+1}(x) = p_n(x) + \alpha s_{n+1}(x) = q_n(x, \alpha).$$

It is clear that $\deg q_{n+1} \leqslant n+1$.

LEMMA 3. $q_{n+1}(x) \geqslant 0$ *for all* $x \in [0, 1]$.

Proof. It will be enough to prove that $p_n + \varepsilon s_{n+1} \geqslant 0$ whenever $0 \leqslant \varepsilon < \alpha$. We write

$$-|p_n(x) + \varepsilon s_{n+1}(x)| = |f(x) - (p_n(x) + \varepsilon s_{n+1}(x))| - f(x)$$
$$\leqslant |f(x) - (p_n(x) + \varepsilon s_{n+1}(x))|.$$

If now $s_{n+1}(x) < 0$, then this last expression is negative, so $p_n(x) + \varepsilon s_{n+1}(x) > 0$. If, on the other hand, $s_{n+1}(x) \geqslant 0$, then $p_n(x) + \varepsilon s_{n+1}(x) \geqslant p_n(x) \geqslant 0$.

LEMMA 4. *If x_j is a good intersection value of (f, p_n), then it is a good intersection value of (f, q_{n+1}).*

Proof. We need only suppose that x_j is a crossover value of (f, p_n), because the end intersection case is trivial. If $\alpha = 0$, there is nothing to prove. If $\alpha > 0$ and if x_j is a crossover value of (f, p_n), then choose $0 \leqslant \varepsilon < \alpha$ so that $[f(x) - (p_n(x) + \varepsilon s_{n+1}(x))] \leqslant 0$ when $x_{j-1} < x < x_j$ and $[f(x) - (p_n(x) + \varepsilon s_{n+1}(x))] \geqslant 0$ when $x_j < x < x_{j+1}$, or vice versa. (If $j = 0$, we let $x_{-1} = 0$ and if $j = n$, we let $x_{n+1} = 1$, as a convention.) Now on letting $\varepsilon \to \alpha-$, we have the result, since we may choose y_{j-1} and y_{j+1} so that $f - q_{n+1}$ has no zero in (y_{j-1}, x_j) and no zero in (x_j, y_{j+1}).

The following lemmas, whose geometrical content seems evident, helps illuminate the situation if (f, q_{n+1}) has no good intersection values that are not among the good intersection values of (f, p_n). (We think of $\Delta(x)$ as $f(x) - q_{n+1}(x)$, and change the interval $[x_{j-1}, x_j]$ into $[0, 1]$.)

DEFINITION. Let $f(x)$ and $g(x)$ be two continuous functions on $[0, 1]$. Suppose that $f(0) = g(0) = 0$ and that $g(x) \leqslant f(x)$ for all $x \in [0, 1]$. We say that the graphs of f and g are tangent at 0 provided that

$$\liminf_{x \to 0+} \frac{f(x) - g(x)}{x} = 0.$$

A similar definition holds for tangency at $x = 1$.

LEMMA 5. *Let $\Delta(x)$ be a continuous function on $[0, 1]$, say, such that $\Delta(0) = \Delta(1) = 0$ and $\Delta(x) > 0$ for all $x \in (0, 1)$. Let $s(x)$ be a polynomial such that $s(0) = s(1) = 0$, $s'(0) \neq 0$, $s'(1) \neq 0$, and $s(x) > 0$ for all $x \in (0, 1)$. If $s(x) < \Delta(x)$ for all $x \in (0, 1)$ and if the graphs of s and Δ are not tangent at either $x = 0$ or $x = 1$, then there exists an $\varepsilon > 0$ such that $(1 + \varepsilon)$ $s(x) \leqslant \Delta(x)$ for all $x \in [0, 1]$.*

Proof. Otherwise, for $\varepsilon = 1/n$, there exists x_n in $(0, 1)$ so that $(1 + (1/n))$ $s(x_n) \geqslant \Delta(x_n)$. Choose a convergent subsequence $x_n \to x_0$. Since $s(x) < \Delta(x)$

in the open interval, x_0 must be an endpoint, say $x_0 = 0$. Now $\Delta(x_n) - s(x_n) \leqslant (1/n)\, s(x_n)$ so that

$$0 \leqslant \frac{\Delta(x_n) - s(x_n)}{x_n} \leqslant \frac{1}{n} \frac{s(x_n)}{x_n}.$$

Hence

$$\liminf_{x \to 0+} \frac{\Delta(x) - s(x)}{x} = 0$$

since $s(x_n)/x_n \to s'(0)$, and the tangency is proved.

LEMMA 6. *Suppose instead that* $\Delta(0) = s(0) = 0$, $s'(0) > 0$, *that* $s(x) < \Delta(x)$ *for all* $x \in (0, 1]$, *and that the graphs of* $s(x)$ *and* $\Delta(x)$ *are not tangent at* $x = 0$. *Then the same conclusion holds.*

Proof. Easier than of Lemma 5, and hence omitted.

LEMMA 7. *Suppose* $\Delta(x) \geqslant 0$ *for all* $x \in [0, 1]$, *that* $\Delta(x) > 0$ *for all* $x \in (0, 1)$, *and that for each* $\varepsilon > 0$, $\varepsilon s(x) \leqslant \Delta(x)$ *fails for some* $x = x_\varepsilon$ *in* $[0, 1]$. *Then the graphs of* $\Delta(x)$ *and* 0 *are tangent at one of the endpoints.*

Proof. Otherwise, for each n, there exists a point x_n such that $\Delta(x_n) \leqslant s(x_n)/n$. Passing to a subsequence, we may suppose that $x_n \to 0$ ($x_n \to 1$ is handled similarly.) Then

$$\frac{\Delta(x_n)}{x_n} \leqslant \frac{1}{n} \frac{s(x_n)}{x_n}$$

and $\liminf \Delta(x)/x = 0$, as was to be proved.

We can now complete the proof of the Theorem. There are a number of possible cases (not entirely disjoint) that we shall be concerned with. It might be good to think of our proof as a computer program. In Cases A, B, and C, the program produces p_{n+1} and stops. In Cases D, D′, E, E′, and F, F′, there is possible trouble at an endpoint. In each of these cases, we might proceed next to Case C, produce p_{n+1}, and then stop. If we don't do this, then we go this time again to some one of the Cases D, D′, E, E′, or F, F′, but after this, we must surely go to Case C, produce p_{n+1}, and then stop.

Case A. q_{n+1} has a zero in $(0, 1)$.

Case B. q_{n+1} has no zero in $(0, 1)$ and (f, q_{n+1}) has a good intersection value that is not a good intersection value of (f, p_n).

Case C. q_{n+1} has no zero in $[0, 1]$ and every good intersection value of (f, q_{n+1}) is a good intersection value of (f, p_n).

Case D. q_{n+1} has no zero in $(0, 1)$ but $q_{n+1}(0) = 0$, $q'_{n+1}(0) \neq 0$ and $q_{n+1}(x) > f(x)$ for all $x \in (0, h)$ for some $h > 0$.

Case D′. Same as Case D, but with $x = 0$ replaced by $x = 1$ throughout.

Case E. Same as Case D but with $q'_{n+1}(0) = 0$.

Case E′. Same as Case D′ but with $q'_{n+1}(1) = 0$.

Case F. Same as Case D, but with $q_{n+1}(x) < f(x)$ for all $x \in (0, h)$.

Case F′. Same as Case D′ but with $q_{n+1}(x) < f(x)$ for all $x \in (1 - h, 1)$.

Treatment of case A. (For some $t \in (0, 1)$, $q_{n+1}(t) = 0$.) In this case we must have $\alpha > 0$. Since $p_n(t) > 0$ and since $q_{n+1}(t) = 0$, we have $q_{n+1}(t) < p_n(t)$, and thus $s_{n+1}(t) < 0$. In particular, $q_{n+1}(x) < p_n(x)$ over a neighborhood of t. Also, $f(t) < p_n(t)$. We claim that $f(t) = 0$. For if $f(t) > 0$, then for ε a little smaller than α, but still positive, we would have $f(t) - p_n(t) < 0$, $f(t) - q_{n+1}(t) > 0$, and hence $|f(t) - (p_n(t) + \varepsilon s_{n+1}(t))| > 0$, which would contradict the definition of α.

We also see that $q_{n+1}(x) \geqslant f(x)$ over a deleted neighborhood of t, since $s_{n+1}(x)$ does not change sign near t. Now unless $q_{n+1} = f$ at infinitely many points (in which case we choose $p_{n+1} = q_{n+1}$), q_{n+1} must be strictly greater than f over a deleted neighborhood of t.

We now choose

$$p^*_{n+1}(x) = q_{n+1}\left(\frac{x}{1+\delta}\right),$$

where we will choose δ very small and positive to satisfy a number of requirements. First of all, we may by Lemma 2 be sure that (f, p^*_{n+1}) has a corresponding cossover value near to each crossover value of (f, q_{n+1}). If 0 is an intersection value of (f, q_{n+1}), then it is an intersection value of (f, p^*_{n+1}). This is not necessarily so for the right-hand endpoint $x = 1$, however. Let us further restrict δ so that $f(t + \delta t) > 0$. We shall need two more restrictions on the size of δ.

We have $x_k < t < x_{k+1}$, where x_k and x_{k+1} are either good intersection values of (f, q_n) or else $x_{k+1} = 1$ and $q_{n+1}(1) > f(1)$ or else $x_k = 0$ and $q_{n+1}(0) > f(0)$, by convention on the labelling x_{-1} and x_{n+1}. Choose t' with $x_k < t' < t$ and t'' with $t < t'' < x_{k+1}$. Choose $\varepsilon > 0$ so that $q_{n+1}(t') > f(t') + 2\varepsilon$ and $q_{n+1}(t'') > f(t'') + 2\varepsilon$. Now we choose δ so small that $|q_{n+1}(x) - q_{n+1}(x/(1 + \delta))| < \varepsilon$ for all $x \in [0, 1]$. Now $p^*_{n+1}(t') > f(t')$

and $p_{n+1}^*(t+\delta t)=0<f(t+\delta t)$. Hence there is at least one crossover value of (f, p_{n+1}^*) between x_k and $t+\delta t$. Similarly, there is at least one crossover value of (f, p_{n+1}^*) between $t+\delta t$ and x_{k+1}. This implies that there are at least two more crossover values of (f, p_{n+1}^*) than there are of (f, p_n). On the other hand, if 1 is an end intersection value of (f, p_n), it need not be one of (f, p_{n+1}^*). But by taking a sure gain of at least 2 and a possible loss of 1, we have at least $n+2$ good intersection values of (f, p_{n+1}^*). Since we have no guarantee that p_{n+1}^* is strictly positive over $(0,1)$, the final step is to let $p_{n+1}(x)=p_{n+1}^*(x)+\lambda x$, where $\lambda>0$ is chosen so small that, by Lemma 2, (f, p_{n+1}) has at least as many crossover values as has (f, p_{n+1}^*). Since $p_{n+1}^*(0)=p_{n+1}(0)$, and since $p_{n+1}(x)>0$ for all $x\in(0,1)$, the polynomial p_{n+1} satisfies all the requirements.

Treatment of Case B. (q_{n+1} has no zero in $(0,1)$ and (f, q_{n+1}) has a good intersection value not among the $\{x_i\}$.) Merely choose $p_{n+1}=q_{n+1}$.

Treatment of Case C. (All the intersection values of (f, q_{n+1}) are among the $\{x_i\}$ and $q_{n+1}>0$ over $[0,1]$.) Let $p_{n+1}^*=q_{n+1}+\delta s_{n+1}$, where we will choose $\delta>0$ so small that by Lemma 2, (f, p_{n+1}^*) has a corresponding crossover value near each crossover value of (f, q_{n+1}). Also, we restrict δ to be so small that $q_{n+1}+\delta s_{n+1}>0$ over $[0,1]$. Since $q_{n+1}+\delta s_{n+1}=p_n+(\alpha+\delta)s_{n+1}$, where $\alpha+\delta>\alpha$, we see by the definition of α that (f, p_{n+1}^*) must have an intersection value x' that is not among the x_i. If it is a good intersection value, then we simply let $p_{n+1}=p_{n+1}^*$. Otherwise it is a kiss value, and by changing δ slightly, we can push the kiss point past the graph of f so that it splits into two crossover points.

Remark. The purpose of our treatment of Cases D–F′ is to handle tangencies at endpoints. We do this by modifying the slope of q_{n+1} near the offending endpoint. This might lose an endpoint intersection value, but compensates by gaining new crossover values. The net result is a reduction to cases that have already been treated.

Treatment of Case D. We let $q_{n+1}^*(x)=q_n(x/1+\delta)$, where we choose $\delta>0$ so small that $(f, {}_{n+1}^*)$ has corresponding crossover values near each crossover value of (f, q_{n+1}). If $x=1$ was an intersection value of (f, q_{n+1}), it is not necessarily an intersection value of (f, q_{n+1}^*). But in any event, there must exist a crossover point x' of (f, q_{n+1}^*) with $0=x_0<x'<x_1$. The method of proof is like that of Lemma 5, so we omit it. Note that the net effect is that we are now in Case F, and $x=1$ is definitely not an intersection value of (x, q_{n+1}^*).

Treatment of Case D′. The same as that of Case D, with $x=0$ replaced by $x=1$.

Treatment of Case E'. Let $p^*_{n+1}(x) = q_{n+1}(x) + \delta(1-x)^2$ for $\delta > 0$ so small that (f, p^*_{n+1}) has a corresponding crossover value near each crossover value of (f, q_{n+1}). Note that $q''_{n+1}(1) \geqslant 0$ since q_{n+1} would otherwise be negative for x slightly smaller than 1. Let $p_{n+1}(x) = p^*_{n+1}(x + \varepsilon x)$, where $\varepsilon > 0$ is chosen with an eye to corresponding crossover points with (f, p^*_{n+1}). Further restrict ε so that $f(1/(1+\varepsilon)) > 0$. It is easy to see that (f, p^*_{n+1}) has two crossover values in the interval (x_{n-1}, x_n), where $x_n = 1$. However, neither 0 nor 1 is an intersection value of (f, p^*_{n+1}). But then (f, p^*_{n+1}) still has at least $n+1$ good intersection values, and we can move to Case C to construct p_{n-1}.

Treatment of Case E. The same as that of Case E', but with $x = 1$ replaced by $x = 0$ throughout.

Treatment of Case F. We will be brief. Let $q^*_{n+1}(x) = q_{n+1}(x) + \delta(1-x)$, where δ is small. One loses the endpoint $x = 0$ and gains a crossover value near $x = 0$.

Treatment of Case F'. Essentially the same.

On tracing the "program" described above, one arrives at Case A, B, or C within at most two steps. In each of these three cases, one constructs p_{n+1}, and the proof is done.

Remarks. We outline an alternative approach to a proof of our Theorem, suggested to us by G. Halász, that might be simpler than ours. However, writing down a detailed proof along these lines might prove troublesome. For simplicity, we take $n = 2m$ even. Let $q_n(x) = \lambda(x-t_1)^2 (x-t_2)^2 \cdots (x-t_m)^2$, where the t_i are chosen in (0, 1). For λ large and positive, (f, q_n) will have n intersection values. Then one uses the "blowing up" argument with the s_n as before to create the needed extra intersection value.

There is a closely related problem whether, if f is infinitely differentiable, we may coalesce all the $n+1$ points of interpolation into one single point of multiplicity $n+1$. This means that p_n is the nth section of the Taylor series for f at some point x_0 in $[0, 1]$. The following argument by G. Halász, whom we thank for his help, shows that this is not possible in general.

Take n large and odd, and let $f(x)$ be a non-negative polynomial of degree $n+1$ such that $f^{(n+1)}(x) = c$ for all $x \in [0, 1]$, where c is a positive constant. It is possible to do this so that $f(0) = f(1) = 0$. Now if p_n is the nth section of the Taylor series for f at x_0 then clearly

$$f(x) = p_n(x) + \frac{c(x-x_0)^{n+1}}{(n+1)!}.$$

Then surely, since $n+1$ is even, either $f(0) > p_n(0)$ or $f(1) > p_n(1)$, which rules out p_n being non-negative throughout $[0, 1]$, since $f(0) = f(1) = 0$. A simple argument shows that if n is even, and if we choose p_n so that $p_n^{(n)}(x) \equiv M$, where $M = \max f^{(n)} = f^{(n)}(x_0)$, then $p_n \geqslant f$, so the answer is affirmative if n is even.

Bruce Reznick has pointed out that for $f(x) = x^3$ on $[0, 1]$, it is impossible to interpolate at 0, a, and 1, if $0 < a < 1$, with a non negative polynomial p of degree $\leqslant 2$. He has also suggested the interesting problem whether our theorem remains true if one drops the hypothesis that f be continuous. It seems not to be easy even in the case $n = 2$.