F. W. Luttmann*
T. J. Rivlin

Some Numerical Experiments in the Theory of Polynomial Interpolation

Abstract: An important unsolved problem in the theory of polynomial interpolation is that of finding a set of nodes which is optimal in the sense that it leads to minimal Lebesgue constants. In this paper results connected to this problem are obtained, and some conjectures are presented based upon numerical evidence garnered from extensive computations.

Introduction

• *General*

If the values of a function, $f(x)$, are known for a finite set of x values in a given interval, then a polynomial which takes on the same values at these x values offers a particularly simple analytic approximation to $f(x)$ throughout the interval. This approximating technique is called polynomial interpolation. Its effectiveness depends on the smoothness of the function being sampled (if the function is unknown, some hypothetical smoothness must be chosen), on the number and choice of points at which the function is sampled, and on what measure is used to determine how far the interpolating polynomial is from the given function. In this paper we take as measure of the error of approximation the greatest vertical distance between the graph of the function and that of the interpolating polynomial over the entire interval under consideration. We consider the problem of how to choose n points at which to sample any function of given smoothness in order that the error should be as small as possible. In general, the location of optimal interpolation points is unknown and we have made some numerical computations which may indicate the direction in which such points should be sought.

* University of Arizona, Tucson, Arizona.

• *Particular*

For each $n = 1, 2, \cdots$ let $x_1^{(n)}, \cdots, x_n^{(n)}$ be real numbers which we call "nodes", satisfying

$$-1 \leq x_n^{(n)} < x_{n-1}^{(n)} < \cdots < x_1^{(n)} \leq 1.$$

The infinite triangular matrix of nodes

$$x_1^{(1)}, x_1^{(2)} x_2^{(2)}, \cdots, x_1^{(n)} \cdots x_n^{(n)}, \cdots$$

is denoted by X. Let $f(x)$ be a function defined on I: $-1 \leq x \leq 1$. Then there exists a unique polynomial of degree not exceeding $n - 1$ which agrees with $f(x)$ at $x_1^{(n)}, \cdots, x_n^{(n)}$ for each $n = 1, 2, \cdots$. Given n we obtain this polynomial by first forming the polynomials of degree $n - 1$

$$l_i(x) = l_i^{(n)}(x; X) = \frac{\prod_{\substack{j=1 \\ j \neq i}}^{n} (x - x_j^{(n)})}{\prod_{\substack{j=1 \\ j \neq i}}^{n} (x_i^{(n)} - x_j^{(n)})}, \qquad i = 1, \cdots, n. \qquad (1)$$

(When n and X are fixed, and no ambiguity will arise, we drop the superscripts (n) and the node-identification X.) It is clear from (1) that

$$l_i(x_j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} . \tag{2}$$

Now we put

$$L_n(f, X; x) = \sum_{i=1}^{n} f(x_i) l_i(x). \tag{3}$$

Then $L_n(x) = L_n(f, X; x)$ is a polynomial of degree at most $n - 1$ with the property that $L_n(x_i) = f(x_i)$, $i = 1, \cdots, n$. (Moreover, if $Q(x_i) = f(x_i)$, $i = 1, \cdots, n$ and $Q(x)$ is a polynomial of degree not exceeding $n - 1$, then $Q(x) - L_n(x) = 0$ for $x = x_1, \cdots, x_n$ and so $Q(x) \equiv L_n(x)$, thus establishing the above-mentioned uniqueness.) The polynomial $L_n(x)$ is called the *Lagrange interpolating polynomial* of degree $n - 1$ to $f(x)$ at $x_1, \cdots, x_n$.

It is rather natural to expect, *a priori*, that the sequence $L_1(f, X; x)$, $L_2(f, X; x)$, $\cdots$, $L_n(f, X; x)$, $\cdots$ should converge uniformly to $f(x)$ for $x \in I$, at least for a continuous $f(x)$. That this expectation is illusory is the result of the following theorem of G. Faber [1].

Theorem. Given X there exists a continuous function on I, $f(x)$, such that $\{L_n(f, X; x)\}$ does not converge uniformly on I.

In contrast to this, if $f(z)$ is an entire function then for any X, $L_n(f, X; x)$ converges uniformly to $f(x)$ in I.

The convergence properties of the Lagrange interpolating polynomials are closely connected to the behavior of the Lebesgue functions

$$\lambda_n(x) = \lambda_n(X, x) = \sum_{i=1}^{n} |l_i(x)|, \tag{4}$$

and the Lebesgue constants

$$\Lambda_n = \Lambda_n(X) = \max_{-1 \le x \le 1} \lambda_n(x). \tag{5}$$

In general, the smaller the Λ_n the better the convergence (but see Erdös and Turan [2]). Erdös [3] has shown that there exists a positive constant, C_1, such that for any X

$$\Lambda_n > (2/\pi) \log n - C_1. \tag{6}$$

Moreover, according to Erdös [3], if we denote the matrix of nodes whose n^{th} row is $\cos(\pi/2n)$, $\cos(3\pi/2n)$, $\cdots$, $\cos((2n - 1)\pi/2n)$ (i.e., the zeros of the Chebyshev polynomial of degree n, $T_n(x)$ where $T_n(x) = \cos n\theta$, $x = \cos\theta$, $0 \le \theta \le \pi$) by T then there is a positive constant C_2 such that

$$\Lambda_n(T) < (2/\pi) \log n + C_2. \tag{7}$$

It is clear from Eq. (6) and Eq. (7) that the matrix of nodes T is an excellent choice for interpolation, as far as convergence properties are concerned. However, it can be shown that there exists at least one matrix of nodes X^* such that $\Lambda_n^* = \Lambda_n(X^*) \le \Lambda_n(X)$, $n = 1, 2, \cdots$ for any matrix of nodes X. We shall call X^* an *extremal* matrix of nodes. The problem of finding an extremal matrix of nodes is a famous unsolved problem of interpolation theory. It is clear from Eq. (6) and Eq. (7) that

$$\Lambda_n^* - \frac{2}{\pi} \log n = 0(1), \; n \to \infty.$$

In this paper we wish to present some conjectures based upon automatic digital computation, and some proved results as well, in this circle of ideas. (Background material mentioned above for which no reference is given can be found in Natanson [4].)

Some general results

We want first to collect some elementary properties of the Lebesgue functions. It is clear from Equation (4) that the Lebesgue functions, $\lambda_n(x)$, are *piecewise polynomials.* More precisely, if $I(k, n)$ denotes the interval $[x_k^{(n)}, x_{k-1}^{(n)}]$, $k = 2, 3, \cdots, n$ while $I(1, n)$ is $(x_1^{(n)}, \infty)$ and $I(n + 1, n)$ is $(-\infty, x_n^{(n)})$, then $\lambda_n(x)$ is a polynomial of degree at most $n - 1$ in each interval $I(k, n)$, $k = 1, \cdots, n + 1$. Let $\lambda_n^k(x)$ be the polynomial which coincides with $\lambda_n(x)$ on $I(k, n)$. The results given below as Assertions (A-1)–(A-5) are easy consequences of the definitions of Eqs. (1) and (4).

(A-1) If $x \in I(k, n)$ then for $i = 1, \cdots, n$,

$$\operatorname{sgn} l_i(x) = \begin{cases} (-1)^{i-k}, & i \ge k \\ (-1)^{k-i-1}, & i < k. \end{cases}$$

(A-2) $$\lambda_n^1(x_j) = \sum_{i=1}^{n} (-1)^{i-1} l_i(x_j) = (-1)^{j-1}, \quad j = 1, \cdots, n$$

and

$$\lambda_n^{n+1}(x_j) = \sum_{i=1}^{n} (-1)^{n-i} l_i(x_j) = (-1)^{n-j}, \quad j = 1, \cdots, n.$$

Hence $\lambda_n^1(x)$ and $\lambda_n^{n+1}(x)$ each have exactly one simple zero in each interval $I(k, n)$, $k = 2, 3, \cdots, n$, and no other zeros. Consequently all derivatives of $\lambda_n^1(x)$ and $\lambda_n^{n+1}(x)$ (which are not identically zero) have all their zeros in (x_n, x_1).

(A-3) $\lambda_n^1(x)$ is strictly increasing and convex in $x \ge x_1^{(n)}$, and so the same is true of $\lambda_n(x)$.
Also, $\lambda_n^{n+1}(x)$ is strictly decreasing and convex in $x \le x_n^{(n)}$, and so the same is true of $\lambda_n(x)$.

(A-4) Since $1 = L_n(1, X; x)$

$$= \sum_{i=1}^{n} l_i(x) \leq \sum_{i=1}^{n} |l_i(x)| = \lambda_n(x),$$

we see that

$$\lambda_n(x) \geq 1. \tag{8}$$

Moreover, equality occurs in Eq. (8) for $x = x_i^{(n)}$, $i = 1, \cdots, n$, and for $n \geq 3$ only for these values of x.

(A-5) If $n \geq 3$ there is exactly one point in $I(k, n)$ at which $\lambda_n^k(x)$, and hence $\lambda_n(x)$, takes on its maximum value on $I(k, n)$.

To prove (A-5) it suffices to show, in view of (A-4), that the derivative of $\lambda_n^k(x)$ has at most one zero in $I(k, n)$. Suppose this to be false for some k among $1, \cdots, n$, then the derivative of $\lambda_n^k(x)$ has three zeros in $I(k, n)$ since $\lambda_n^k(x) = 1$ at the end points of $I(k, n)$ and $\lambda_n^k(x) \geq 1$ throughout $I(k, n)$. But it is a consequence of (A-1) that $\lambda_n^k(x)$ has at least $(k - 2)$ sign changes in (x_{k-1}, x_1) and at least $(n - k)$ sign changes in (x_n, x_k). Therefore the derivative of $\lambda_n^k(x)$ has at least $(k - 3)$ zeros in (x_{k-1}, x_1), at least $(n - k - 1)$ zeros in (x_n, x_k), and three zeros in (x_k, x_{k-1}), hence a total of at least $(n - 1)$ zeros for a polynomial of degree at most $n - 2$. This leads to a contradiction since for $n \geq 3$, $\lambda_n^k(x) \not\equiv 1$, and (A-5) must be true.

In summary then, for $n \geq 3$ the $\lambda_n(x)$ are piecewise polynomials satisfying Eq. (8), taking the value 1 only at $x = x_i$, $i = 1, \cdots, n$, having a single maximum between consecutive nodes, monotone decreasing and convex in $(-1, x_n)$ and monotone increasing and convex in $(x_1, 1)$.

Next, suppose

$$M_n(X) = \max_{x_n \leq x \leq x_1} \lambda_n(X, x).$$

And suppose

$$y_i = ax_i + b, \; i = 1, \cdots, n,$$

where $a > 0$ and a, b are chosen so that $y_i \, \varepsilon \, I$, $i = 1, \cdots, n$. Such ordered pairs, (a, b), will be called "admissible" (with respect to X).

Theorem 1. For an extremal matrix of nodes, X^*, $\Lambda_n(X^*) = M_n(X^*)$, $n = 1, 2, \cdots$.

Proof. Suppose the theorem were false. Then $M_n(X^*) < \Lambda_n(X^*)$ for some n. Let $x_1^*, \cdots, x_n^*$ be the n^{th} row of X^*. Choose a, b to satisfy $ax_1^* + b = 1$, $ax_n^* + b = -1$. (a, b) is admissible. Put $ax_i^* + b = y_i$, $i = 1, \cdots, n$ and let Y denote a matrix of nodes whose n^{th} row is $y_1, \cdots, y_n$. Then

$$\Lambda_n(Y) = \max_{-1 \leq y \leq 1} \sum_i \prod_j \left| \frac{y - y_j}{y_i - y_j} \right|$$

$$= \max_{x_n^* \leq x \leq x_1^*} \sum_i \prod_j \left| \frac{(ax + b) - (ax_j^* + b)}{(ax_i^* + b) - (ax_j^* + b)} \right|$$

$$= M_n(X^*) < \Lambda_n(X^*),$$

contradicting the extremality of X^*. *The theorem is proved.*

By (A-3) there exist $\zeta \geq 1$ and $\eta \leq -1$ such that

$$\lambda_n(X, \zeta) = \lambda_n(X, \eta) = \Lambda_n(X).$$

Let a, b satisfy $a\zeta + b = 1$, $a\eta + b = -1$. (a, b) is admissible. If $ax_i + b = z_i$, $i = 1, \cdots, n$, and Z is a matrix of nodes whose n^{th} row is $z_1, \cdots, z_n$ then

$$\Lambda_n(Z) = \Lambda_n(X)$$

and

$$\lambda_n(Z, \pm 1) = \Lambda_n(Z).$$

Note that ± 1 are not included among $z_1, \cdots, z_n$ $(n \geq 3)$.

Suppose $\Lambda_n(X) = M_n(X)$. It is clear that $M_n(X) = M_n(Z)$, hence $M_n(Z) = \Lambda_n(Z)$. Now, any (a, b) satisfying both

$$a > 1, \; |b| \leq a - 1$$

and

$$az_1 + b \leq 1$$

$$az_n + b \geq -1,$$

is admissible, and there exist infinitely many such (a, b).

Choose one such (a, b), put $w_i = az_i + b$, $i = 1, \cdots, n$, and let W be a matrix of nodes whose n^{th} row is $w_1, \cdots, w_n$, then $W \neq Z$ and $\Lambda_n(W) = \Lambda_n(Z)$.

In the course of this discussion we have proved, in view of Theorem 1,

Theorem 2. An extremal set of nodes, X^*, is not unique.

Some numerical results

In view of Eqs. (6) and (7) it seems worthwhile to investigate the Chebyshev nodes, T, further. The IBM 7094 was used to explore the behavior of $\lambda_n(T, x)$ for $n = 3, \cdots, 40$. Since each row of T is symmetric with respect to the origin, $\lambda_n(T, x)$ is an even function of x and so its behavior on I is determined by its behavior on $[-1, 0]$. We calculated $\lambda_n(T, -1)$ and the (single) maximum of $\lambda_n(T, x)$ on each subinterval between consecutive nodes that contains points of $[-1, 0]$. The maximum on a subinterval was calculated by evaluating $\lambda_n(T, x)$ at points of successive bisection of the subinterval. After each pass the largest value of $\lambda_n(T, x)$ on the points of the preceding subdivision was stored, and the process was repeated on a bisected mesh until the difference between two consecutive largest values became less than 10^{-5}. The current largest value of $\lambda_n(T, x)$ was then taken to be the maximum of $\lambda_n(T, x)$ on the subinterval under consideration.

The computations suggest that $\Lambda_n(T) = \lambda_n(T, \pm 1)$ and that $\lambda_n(T, -1)$ is substantially larger than the maxima on the subintervals. In addition the maxima are strictly decreasing as we proceed into the interval $[-1, 0]$ from the left-hand end point.

We are thus led to

Conjecture 1. $\lambda_n(T, x) < \lambda_n(T, 1)$ *for* $-1 < x < 1$.

This conjecture focuses attention on $\lambda_n(T, 1)$ as a function of n. We proceed next to determine the asymptotic behavior of this function as $n \to \infty$.

If we put

$$\theta_j^{(n)} = (2j - 1)\pi/2n,\ j = 1, \cdots, n, \tag{9}$$

and

$$x_j^{(n)} = \cos \theta_j^{(n)},\ j = 1, \cdots, n;\ n = 1, 2, \cdots, \tag{10}$$

the resulting matrix of nodes is T. Then

$$\lambda_n(x) = \lambda_n(T, x) = \frac{1}{n} \sum_{j=1}^{n} \frac{|\cos n\theta|}{|\cos \theta - \cos \theta_j|} \sin \theta_j \tag{11}$$

where

$$x = \cos \theta,\ 0 \leq \theta \leq \pi. \tag{12}$$

Theorem 3:

$$\lim_{n\to\infty} \left[\lambda_n(1) - \left(\frac{2}{\pi} \log n + \frac{2}{\pi} \left(\log \frac{8}{\pi} + \gamma \right) \right) \right] = 0. \tag{13}$$

(γ is Euler's constant, $\gamma = .5772 \cdots$).

Proof. In view of Eq. (11) we have

$$\lambda_n(1) = \frac{1}{n} \sum_{j=1}^{n} \frac{\sin \theta_j}{1 - \cos \theta_j} = \frac{1}{n} \sum_{j=1}^{n} \cot \frac{\theta_j}{2}.$$

Therefore,

$$\pi\lambda_n(1) = \frac{\pi}{n} \sum_{j=1}^{n} \left(\cot \frac{\theta_j}{2} - \frac{1}{\theta_j/2} \right) + \frac{\pi}{n} \sum_{j=1}^{n} \frac{1}{\theta_j/2}.$$

But

$$\lim_{n\to\infty} \frac{\pi}{n} \sum_{j=1}^{n} \left(\cot \frac{\theta_j}{2} - \frac{1}{\theta_j/2} \right) = \int_0^{\pi} \left(\cot \frac{\theta}{2} - \frac{1}{\theta/2} \right) d\theta$$

$$= 2 \log \frac{2}{\pi},$$

while since

$$\frac{2\pi}{n} \sum_{j=1}^{n} \frac{1}{\theta_j} = 4 \sum_{j=1}^{n} \frac{1}{2j - 1} = 4 \left(\sum_{j=1}^{2n} \frac{1}{j} - \frac{1}{2} \sum_{j=1}^{n} \frac{1}{j} \right)$$

we obtain

$$\lim_{n\to\infty} \left(\frac{2\pi}{n} \sum_{j=1}^{n} \frac{1}{\theta_j} - 4 \log 2n + 2 \log n \right)$$

$$= 4(\gamma - \tfrac{1}{2}\gamma) = 2\gamma.$$

The theorem is established.

A consequence of Theorem 3 is that the constant C_2 of Eq. (7) must satisfy

$$C_2 \geq \frac{2}{\pi} \left(\log \frac{8}{\pi} + \gamma \right) = .9625 \cdots.$$

If we assume that Conjecture 1 is true we obtain interesting results:

Theorem 4. If we assume Conjecture 1 then $\Lambda_n(X^*) < \Lambda_n(T)$ *for* $n \geq 2$.†

Proof. Conjecture 1 implies that $M_n(T) < \Lambda_n(T)$ and Theorem 4 then follows from Theorem 1.

Moreover, if $x_j^{(n)}$, $j = 1, \cdots, n$ is defined by Eq. (10) and we put

$$e_j^{(n)} = ax_j^{(n)} + b,\ j = 1, \cdots, n,$$

where (a, b) is the (necessarily admissible) solution of

$$ax_1^{(n)} + b = 1$$

$$ax_n^{(n)} + b = -1,$$

we obtain a new matrix of nodes, E, the "expanded" Chebyshev nodes. (The Chebyshev nodes have been expanded as far as possible.) We obtain easily

Theorem 5. If we assume Conjecture 1 then $\Lambda_n(E) < \Lambda_n(T)$.

Proof. $\Lambda_n(E) = M_n(T)$ and Conjecture 1 implies that $M_n(T) < \Lambda_n(T)$.

The Fekete nodes

The problem of finding an extremal matrix of nodes, X^*, is that of determining an X^* that minimizes

$$\Lambda_n(X) = \max_{-1\leq x\leq 1} \sum |l_i^{(n)}(x; X)|$$

among all choices of X. A similar problem, that of

† $\Lambda_2(X^*) < \Lambda_2(T)$ is easy to establish without Conjecture 1. Also S. N. Bernstein has shown, without Conjecture 1, that $\Lambda_3(X^*) < \Lambda_3(T)$, in *Sur la limitation des valeurs d'un polynome $P_n(x)$ de degré n sur tout un segment par ses valeurs en n + 1 points du segment*. Izvestia Akad. Nauk S.S.S.R., Classe des Sciences Math. et Naturelles, 1931, pp. 1025–1050 (See page 1027).

determining a set of nodes that minimizes

$$\Phi_n(X) = \max_{-1 \le x \le 1} \sum [l_i^{(n)}(x;\ X)]^2$$

among all choices of X, was solved by Fejér [5]. An extremal set of nodes is $X = F$, where F is the array of Fekete nodes, which are defined as follows. The n^{th} row of F consists of the zeros of $(x^2 - 1)\, P'_{n-1}(x)$, where $P_n(x)$ is the Legendre polynomial of degree n. It is clear that these zeros all lie in I. (We name these points after Fekete since it can be shown that the n^{th} row of F consists of the n points of I for which

$$\prod_{\substack{i<j \\ i,j=1}}^{n} |x_i - x_j|$$

is greatest among all choices of $x_1, \cdots, x_n \in I$. Such points are intimately connected with Fekete's concept of the "transfinite diameter" of a point set, in this case I.)

Fejér established that $\Phi_n(F) = 1$, which, since $\Phi_n(X) \geq 1$ is obvious for any X, proves the extremality of F. Further, $\Phi_n(F) = 1$ in conjunction with Schwarz' inequality gives

$$\Lambda_n(F) \leq n^{1/2}. \tag{14}$$

Numerical evidence suggests that the bound in Eq. (14) can be substantially improved. The computation of $\Lambda_n(F)$ for $n = 3, \cdots, 40$ was similar to that of $\Lambda_n(T)$ as described earlier (page 189).

Conjecture 2. $\Lambda_n(E) < \Lambda_n(F) < \Lambda_n(T)$, $n > 3$.

Conjecture 3. The relative maxima of $\lambda_n(F, x)$ *on subintervals between nodes increase in* $[-1, 0]$. $\lambda_n(F, x)$ *attains its maximum on* I *at* $x = 0$ *for even* n *and "near"* $x = 0$ *for odd* n.

References

1. Faber, G., "Über die interpolatorische Darstellung stetiger Funktionen," *Jahresbericht der Deutschen Math.* Ver. **23**, 192–210 (1914).
2. Erdös, P., and Turan, P., "On the Role of the Lebesgue Functions in the Theory of Lagrange Interpolation," *Acta Math. Acad. Sci. Hung.* VI, 47–66 (1955).
3. Erdös, P., "Problems and Results on the Theory of Interpolation, II." *Acta Math. Acad. Sci. Hung.* **XII**, 235–244 (1961).
4. Natanson, I. P., *Konstructive Funktionentheorie*, Akademie-Verlag, Berlin (1955).
5. Fejér, L., "Bestimmung derjenigen Abszissen eines Intervalles, für welche die Quadratsumme der Grundfunktionen der Lagrangeschen Interpolation in Intervalle $[-1, 1]$ ein möglichst kleines Maximum besitzt," *Annali della Scuola Norm. Sup. di Pisa* **2**, 1, 3–16 (1932).

Received December 17, 1964.