JOURNAL OF APPROXIMATION THEORY 24, 289–303 (1978)

Proof of the Conjectures of Bernstein and Erdös Concerning the Optimal Nodes for Polynomial Interpolation*

CARL DE BOOR

Mathematics Research Center, University of Wisconsin, Madison, Wisconsin 53706

AND

ALLAN PINKUS

Department of Mathematics, Technion—Israel Institute of Technology, Haifa, Israel

Received April 1, 1977

INTRODUCTION

It is the purpose of this note to complete and extend the work of Kilgore [8] on the optimal nodes in polynomial interpolation.

The problem is as follows. Consider the Banach space $C[a, b]$ of continuous functions on the finite interval $[a, b]$, with the usual norm

$$\|f\| := \max_{a \leqslant x \leqslant b} |f(x)|.$$

Throughout the paper, we take n to be a fixed integer,

$$n \geqslant 2.$$

Corresponding to each point $\mathbf{t}$ in

$$T := \{\mathbf{t} \in \mathbb{R}^{n-1} : a < t_1 < \cdots < t_{n-1} < b\},$$

we construct the linear map $P_{\mathbf{t}}$ of polynomial interpolation in $C[a, b]$ at the $n + 1$ points or nodes $a =: t_0, t_1, \ldots, t_n := b$. In its Lagrange form,

$$P_{\mathbf{t}} f := \sum_{i=0}^{n} f(t_i)\, l_i$$

with

$$l_i(x) := \prod_{j \neq i} \frac{x - t_j}{t_i - t_j}, \qquad i \in [0, n].$$

* Sponsored by the United States Army under Contract No. DAAG29-75-C-0024 and the National Science Foundation under Grant No. MCS75-17385.

0021-9045/78/0244–0289$02.00/0

We want to determine *optimal nodes*, i.e., a point or points $\mathbf{t}^* \in T$ for which

$$\| P_{\mathbf{t}^*} \| = \inf_{\mathbf{t} \in T} \| P_{\mathbf{t}} \|.$$

Here, $\| P_{\mathbf{t}} \| := \sup_{f \in C} \| P_{\mathbf{t}} f \| / \| f \|$, as usual. This problem is motivated by the fact that $P_{\mathbf{t}}$ is a projector on $C[a, b]$ and its range is π_n, the subspace of polynomials of degree $\leqslant n$, which implies that

$$\| f - P_{\mathbf{t}} f \| \leqslant (1 + \| P_{\mathbf{t}} \|) \operatorname{dist}(f, \pi_n).$$

It is well known that $\| P_{\mathbf{t}} \|$ can be computed as

$$\| P_{\mathbf{t}} \| = \| \Lambda_{\mathbf{t}} \|,$$

with

$$\Lambda_{\mathbf{t}} := \sum_{i=0}^{n} | l_i |$$

the *Lebesgue function* of the process. A simple argument shows that $\Lambda_{\mathbf{t}}(x) \geqslant 1$ with equality iff $x \in \{t_0, ..., t_n\}$. Set

$$\lambda_i(\mathbf{t}) := \max_{t_{i-1} \leqslant x \leqslant t_i} \Lambda_{\mathbf{t}}(x) \qquad \text{for} \quad i \in [1, n].$$

In 1931, S. Bernstein [1] conjectured that $\| P_{\mathbf{t}} \|$ is minimal when $\Lambda_{\mathbf{t}}$ *equioscillates*, i.e., when $\lambda_1(\mathbf{t}) = \lambda_2(\mathbf{t}) = \cdots = \lambda_n(\mathbf{t})$. Later, Erdös [7] added to this the conjecture that there is exactly one choice of $\mathbf{t}$ for which $\Lambda_{\mathbf{t}}$ equioscillates and that

$$\min_i \lambda_i(\mathbf{t}) \leqslant \lambda^* := \inf_{\mathbf{s} \in T} \| P_{\mathbf{s}} \| \qquad \text{for every} \quad \mathbf{t} \in T. \tag{1}$$

The latter conjecture appears already in Erdös [6] in the form: "$\min_i \lambda_i(\mathbf{t})$ achieves its maximum when $\Lambda_{\mathbf{t}}$ equioscillates."

Subsequent work on these conjectures and related topics is summarized in Luttmann and Rivlin [11] and in Cheney and Price [4].

Substantial progress in answering these conjectures has come only very recently. Kilgore and Cheney [9] finally showed the existence of $\mathbf{t} \in T$ for which $\Lambda_{\mathbf{t}}$ equioscillates. This result was considerably strengthened by Kilgore [8] who showed that an *optimal* Lebesgue function, i.e., a $\Lambda_{\mathbf{t}}$ for which $\| \Lambda_{\mathbf{t}} \| = \lambda^*$, must necessarily equioscillate.

In the present paper, which is very much based on Kilgore's analysis, we prove the validity of all of the above conjectures. Explicitly, we prove (Theorem 1) that there is only one $\mathbf{t} \in T$ for which $\Lambda_{\mathbf{t}}$ equioscillates, and we prove (Theorem 2) that

$$\lambda_i(\mathbf{t}) \leqslant \lambda_i(\mathbf{s}) \qquad \text{for all } i \in [1, n]$$

cannot hold except in the trivial case when $\mathbf{t} = \mathbf{s}$, from which (1) follows immediately. In addition, we prove analogous results for trigonometric interpolation.

The article is organized as follows. In Section 2, we outline Kilgore's proof of the fact that an optimal Lebesgue function must equioscillate. Section 3 is concerned with the proof of Theorems 1 and 2. In Section 4, we extend these results to the case of trigonometric interpolation. Explicitly, we prove the intuitively obvious fact that trigonometric interpolation on $[0, 2\pi]$ at equidistant nodes is optimal.

2. Kilgore's Result

In this section, we quickly review the proof of Kilgore's result that an optimal Lebesgue function must equioscillate. This we do for completeness and to facilitate its extension to trigonometric interpolation in Section 4. We continue to use the notation introduced in Section 1.

THEOREM (Kilgore [8]). *If* $\|\Lambda_{\mathbf{t}}\| = \lambda^* (= \inf_{\mathbf{t}\in T} \|P_{\mathbf{t}}\|)$, *then* $\Lambda_{\mathbf{t}}$ *equioscillates, i.e., then* $\lambda_1(\mathbf{t}) = \lambda_2(\mathbf{t}) = \cdots = \lambda_n(\mathbf{t})$.

Proof outline. For $i \in [1, n]$, denote by F_i the polynomial of degree $\leqslant n$ which agrees with $\Lambda_{\mathbf{t}}$ on $[t_{i-1}, t_i]$. One easily verifies that F_i is the unique element of π_n for which

$$F_i(t_j) = \begin{cases} (-1)^{i-j+1} & \text{for } j \in [0, i-1] \\ (-1)^{j-i} & \text{for } j \in [i, n]. \end{cases}$$

Furthermore, denote by τ_i the unique point in $[t_{i-1}, t_i]$ at which $\Lambda_{\mathbf{t}}$ and F_i take on the value $\lambda_i(\mathbf{t})$,

$$F_i(\tau_i) = \lambda_i(\mathbf{t}) = \max_{t_{i-1} \leqslant x \leqslant t_i} |F_i(x)| \qquad \text{for all} \quad i \in [1, n].$$

Kilgore points out that the theorem follows at once if it can be shown that

$$\begin{gathered} \text{for all } \mathbf{t} \in T, \text{ all } k \in [1, n], \text{ and all } \boldsymbol{\mu} \text{ close to } \boldsymbol{\lambda}(\mathbf{t}) := (\lambda_i(\mathbf{t}))_1^n, \\ \text{there exists } \mathbf{s} \in T \text{ close to } \mathbf{t} \text{ so that } \lambda_i(\mathbf{s}) = \mu_i \text{ for all } i \neq k. \end{gathered} \tag{2}$$

For, then $\lambda_k(\mathbf{t}) < \|\Lambda_{\mathbf{t}}\|$ for some j implies the existence of $\mathbf{s}$ (near $\mathbf{t}$) for which $\|\Lambda_{\mathbf{s}}\| < \|\Lambda_{\mathbf{t}}\|$.

Kilgore establishes (2) by showing that

$$\text{for } \mathbf{t} \in T, \text{ and } k \in [1, n], \quad J_k := \det(\partial\lambda_i(\mathbf{t})/\partial t_j)_{i=1;\, j=1 \atop i\neq k}^{n \quad n-1} \neq 0. \tag{3}$$

His proof of (3) begins with the observation that

$$\frac{\partial \lambda_i}{\partial t_j} = -F_i'(t_j)\, l_j(\tau_i) = \prod_{k=0}^{n} (\tau_i - t_k) \frac{F_i'(t_j)}{(t_j - \tau_i)} \Big/ \prod_{\substack{k=0\\k\neq j}}^{n} (t_j - t_k)$$

which shows λ_i to be continuously differentiable on T and also shows[1] that (3) is equivalent to

$$\text{for } \mathbf{t} \in T \text{ and } k \in [1, n], \quad \det(q_i(t_j))_{\substack{i=1;j=1\\i\neq k}}^{n\ \ \ n-1} \neq 0, \tag{4}$$

with

$$q_i(x) := F_i'(x)/(x - \tau_i), \qquad i \in [1, n].$$

Since each q_i is a polynomial of degree $\leqslant n-2$, (4) is, in turn, equivalent to the linear independence of any $n-1$ of the n polynomials $q_1, \dots, q_n$. For the proof of this linear independence, Kilgore uses eight lemmas. The first five lemmas lead up to the following

LEMMA 6 of [8]. *On the interval* $[\tau_1, \tau_n]$, *the zeros of* $F_1', \dots, F_n'$ *lie in the pattern*

$$\hat{1}, n, n-1, \dots, 3, \hat{2}, 1, n, n-1, \dots, \hat{3}, 2, 1, n, n-1, \dots,$$
$$3, 2, 1, n, \widehat{n-1}, \dots, 3, 2, 1, \hat{n}.$$

Here, the number i denotes a zero of F_i', *and* $\hat{i}$ *denotes the point* τ_i.

It may be instructive for the reader to consider the following alternative argument which obtains Lemma 6 as an immediate corrollary to the corresponding result for the zeros of $F_1, \dots, F_n$. In this, Fig. 1 may be of help.

For $r \in [1, n]\backslash\{i\}$, F_i changes sign on (t_{r-1}, t_r), hence must have a zero there. Since F_i cannot have more than n zeros, these zeros must all be simple and F_i has no other zeros in $[a, b]$. Let $\sigma_1^{(i)}, \dots, \sigma_{n-1}^{(i)}$ denote these zeros, in increasing order. Then

$$\sigma_r^{(i)} \in \begin{cases} (t_{r-1}, t_r), & \text{for} \quad r < i, \\ (t_r, t_{r+1}), & \text{for} \quad r \geqslant i. \end{cases}$$

If F_i has an additional zero, we denote it by $\sigma_0^{(i)}$ or by $\sigma_n^{(i)}$ depending on whether it is less than a or greater than b, respectively.

LEMMA 1. *For* $i < j$, *the zeros of* F_i *and* F_j *strictly interlace. More precisely,* $\sigma_r^{(j)} < \sigma_r^{(i)}$ *for all applicable r in* $[0, n]$.

[1] We learned only from Kilgore [14] that this elegant and proof simplifying observation is due to Dietrich Braess.

Proof. The function $G_1 := F_i - (-1)^{j-i} F_j$ satisfies

$$G_1(t_k) = \begin{cases} 0 & \text{for } k \in [0, i-1] \cup [j, n] \\ 2(-1)^{k-i} & \text{for } k \in [i, j-1]. \end{cases}$$

Thus, G_1 has at least $i + n + 1 - j$ zeros outside $[t_i, t_{j-1}]$ and $j - 1 - i$ zeros in (t_i, t_{j-1}). Since G_1 is a polynomial of degree $\leqslant n$, it cannot have any additional zeros and all these zeros must be simple. But, since $G_1(t_i) = 2 > 0$, this shows that $(-1)^{i-r} G_1 > 0$ on (t_{r-1}, t_r) for all $r < i$ and so shows that

$$t_{r-1} < \sigma_r^{(j)} < \sigma_r^{(i)} < t_r \qquad \text{for} \quad r \in [1, i-1] \tag{5a}$$

and also

$$\sigma_0^{(j)} < \sigma_0^{(i)} < t_0 \qquad \text{if these exist.} \tag{5b}$$

We have trivially

$$\begin{aligned} t_{i-1} &< \sigma_i^{(j)} < t_i, \\ t_{j-1} &< \sigma_{j-1}^{(i)} < t_j. \end{aligned} \tag{5c}$$

Also, $G_1(t_{j-1}) = 2(-1)^{j-1-i}$, hence $(-1)^{r-i} G_1 > 0$ on (t_r, t_{r+1}) for $r \geqslant j$, and therefore

$$t_r < \sigma_r^{(j)} < \sigma_r^{(i)} < t_{r+1} \qquad \text{for} \quad r \in [j, n-1] \tag{5d}$$

and also

$$t_n < \sigma_n^{(j)} < \sigma_n^{(i)} \qquad \text{if these exist.} \tag{5e}$$

Finally, the function $G_2 := F_i + (-1)^{j-i} F_j$ satisfies

$$G_2(t_k) = \begin{cases} 2(-1)^{k-i-1} & \text{for } k \in [0, i-1] \\ 0 & \text{for } k \in [i, j-1] \\ 2(-1)^{k-i} & \text{for } k \in [j, n]. \end{cases}$$

G_2 has at least the $j - i$ zeros $t_i, \ldots, t_{j-1}$ in $[t_{i-1}, t_j]$ and has at least $i - 1 + n - j$ zeros outside $[t_{i-1}, t_j]$, giving a total of at least $n - 1$ zeros. Since $G_2(t_{i-1}) G_2(t_j) = 4(-1)^{j-i}$, the number of zeros of G_2 in $[t_{i-1}, t_j]$ must be of parity $j - i$. Therefore, since G_2 is of degree $\leqslant n$, it follows that G_2 has no other zeros in $[t_{i-1}, t_j]$. This proves that $(-1)^{r-i} G_2 > 0$ on (t_{r-1}, t_r) for $r \in [i, j]$ and so shows that

$$t_{r-1} < \sigma_{r-1}^{(i)} < \sigma_r^{(j)} < t_r \qquad \text{for} \quad r \in [i+1, j-1]. \tag{5f}$$

Concatenation of (5a–f) proves Lemma 1.

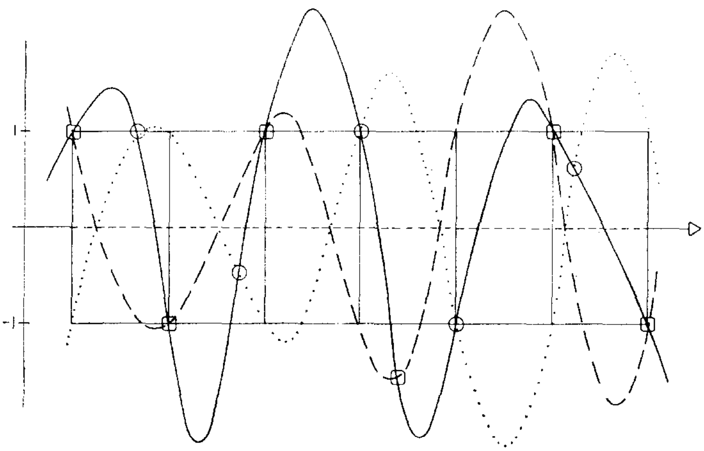

FIG. 1. Schematic drawing of F_i (solid), F_j (dashed) and $-F_j$ (dotted) for $n = 6$, $i = 3$, $j = 5$. The graphs of F_i and $(-1)^{j-i}F_j$ cross at the n points indicated by □, those of F_i and $-(-1)^{j-i} F_j$ cross at the $n-1$ points indicated by ○.

COROLLARY. *The zeros of $F_1, \ldots, F_n$ on $(-\infty, \infty)$ lie in the pattern*

$$\sigma_0^{(I)}, \ldots, \sigma_0^{(1)}, \sigma_1^{(n)}, \ldots, \sigma_1^{(1)}, \sigma_2^{(n)}, \ldots, \sigma_{n-1}^{(1)}, \sigma_n^{(n)}, \ldots, \sigma_n^{(J)},$$

where I and J are certain integers with $1 \leqslant I < J \leqslant n$.

Proof. The corollary is an immediate consequence of Lemma 1 and the additional fact that $\sigma_0^{(1)}$ and $\sigma_n^{(n)}$ necessarily exist.

Since G_1 is of degree n for any i and j, it follows that I equals $J-1$ or $J-2$.

Let now $\tau_r^{(i)}$ denote the zero of F_i' which lies between $\sigma_{r-1}^{(i)}$ and $\sigma_r^{(i)}$. Since the zeros of F_i and F_j interlace for $i \neq j$, V. A. Markov's well-known result [12] implies that the zeros of F_i' and F_j' interlace, and interlace in the same manner. Therefore, the corollary implies

LEMMA 2. *The zeros of $F_1', \ldots, F_n'$ lie in the pattern*

$$\tau_1^{(I)}, \ldots, \tau_1^{(1)}, \tau_2^{(n)}, \ldots, \tau_{n-1}^{(1)}, \tau_n^{(n)}, \ldots, \tau_n^{(J)},$$

where I and J are certain integers with $1 \leqslant I < J \leqslant n$.

Lemma 6 of [8] follows from this since $\tau_i^{(i)} = \tau_i$, all i.

The proof of (4) is now finished as follows. Recall that q_i is a polynomial of degree $\leqslant n-2$ which vanishes at the zeros of F_i' except for τ_i.

We may assume $q_i(\tau_1) > 0$, all i. Lemma 6 then implies that

$$\begin{aligned} \operatorname{sgn} q_i(\tau_j) &= (-1)^{j+1} \qquad \text{for } i, j \in [2, n],\ i \neq j,\\ \operatorname{sgn} q_i(\tau_i) &= (-1)^i \qquad \text{for } i \in [2, n],\\ \operatorname{sgn} q_1(\tau_j) &= (-1)^j \qquad \text{for } j \in [2, n]. \end{aligned}$$

Assume now that $\sum a_k q_k = 0$ for some $\mathbf{a} \neq \mathbf{0}$ with $a_1 \geqslant 0$. Then the set $N := \{k \in [2, n] : a_k < 0\}$ is not empty since $q_k(\tau_1) > 0$ for all k. Set $P := [2, n] \backslash N$ and consider the function

$$f := a_1 q_1 + \sum_{k \in N} a_k q_k = -\sum_{k \in P} a_k q_k .$$

We have

$$(-1)^j f(\tau_j) = \sum_{k \in P} a_k (-1)^{j+1} q_k(\tau_j) \geqslant 0 \qquad \text{for} \quad j \notin P$$

while

$$(-1)^j f(\tau_j) = a_1 (-1)^j q_1(\tau_j) + \sum_{k \in N} (-a_k)(-1)^{j+1} q_k(\tau_j) > 0 \qquad \text{for} \quad j \in P.$$

This shows the polynomial f of degree $\leqslant n-2$ to have $n-1$ weak sign changes, and therefore $f = 0$ and so, in particular, $P = \varnothing$, hence $a_k < 0$ for all $k \in [2, n]$. But since $q_k(\tau_1) > 0$ for all k, it then also follows that $a_1 > 0$.

In summary, $\sum_k a_k q_k = 0$ for some $\mathbf{a} \neq \mathbf{0}$ implies that $a_1 a_k < 0$ for all $k \in [2, n]$. In particular, then $a_k \neq 0$ for all $k \in [1, n]$, and (4) follows.

3. Uniqueness

The central result of this article is the following theorem.

THEOREM 1. *The map $\Gamma : T \to \mathbb{R}^{n-1} : \mathbf{t} \mapsto (\lambda_{i+1}(\mathbf{t}) - \lambda_i(\mathbf{t}))_{i=1}^{n-1}$ is a homeomorphism of T onto $\mathbb{R}^{n-1}$.*

In particular, there is exactly one $\mathbf{t} \in T$ with $\Gamma(\mathbf{t}) = \mathbf{0}$, i.e., exactly one $\mathbf{t}$ for which $\Lambda_\mathbf{t}$ equioscillates. Since Kilgore proved that Γ maps every optimal $\mathbf{t}$ to the point $\mathbf{0} \in \mathbb{R}^{n-1}$, Theorem 1 implies at once the validity of Bernstein's conjecture.

COROLLARY. *If $\Lambda_\mathbf{t}$ equioscillates, then $\| P_\mathbf{t} \| < \| P_\mathbf{s} \|$ for all $\mathbf{s} \neq \mathbf{t}$.*

We use the following two lemmas in the proof of Theorem 1.

LEMMA 3. *The map Γ is a local homeomorphism.*

Proof. It suffices to show that

$$\text{for all } \mathbf{t} \in T,\ \det(\partial(\lambda_{i+1} - \lambda_i)(\mathbf{t})/\partial t_j)_{i,j=1}^{n-1} \neq 0.$$

Expanding this determinant by rows, one obtains

$$\det(\partial(\lambda_{i+1} - \lambda_i)/\partial t_j) = \sum_{k=1}^{n} (-1)^{k+1} J_k$$

where we use again the abbreviation

$$J_k := \det(\partial\lambda_i/\partial t_j)_{\substack{i=1:j=1\\ i\neq k}}^{n\ \ n-1}, \qquad k \in [1, n].$$

Hence, it suffices to show that

$$\text{for some } \epsilon \in \{-1, 1\} \text{ and all } \mathbf{t} \in T,\ k \in [1, n],\ \epsilon(-1)^k J_k(\mathbf{t}) > 0. \tag{6}$$

But, since J_k is a continuous function of $\mathbf{t}$ and never vanishes on T by Kilgore's result, and T is connected, (6) is proved once we show that, for some $\mathbf{t} \in T$,

$$(-1)^k J_k(\mathbf{t})/J_1(\mathbf{t}) < 0 \qquad \text{for } k \in [2, n]. \tag{7}$$

This we could prove by observing that the last part of the argument for Kilgore's Theorem as we gave it in the preceding section gives precise information about the signs of the $(n-1)$-minors of the matrix $(q_i(t_j))$ which is easily translated into the required information about the sign of J_k/J_1, all k. But the following argument is more direct and establishes that

$$\partial\lambda_1/\partial\lambda_k < 0 \qquad \text{for } k \in [2, n], \tag{8}$$

a fact which we need again later.

To prove (7) for some $\mathbf{t}$, observe that, since $J_1(\mathbf{t}) \neq 0$, we can find a continuously differentiable function G on some open neighborhood V of the point $(\lambda_i(\mathbf{t}))_2^n$ and an open neighborhood U of $\mathbf{t}$ so that

$$\lambda_1(\mathbf{s}) = G(\lambda_2(\mathbf{s}),\ldots, \lambda_n(\mathbf{s})) \qquad \text{for all } \mathbf{s} \in U.$$

Also, by Cramer's rule,

$$\partial\lambda_1 = \sum_{k=2}^{n} (-1)^k (J_k/J_1)\, \partial\lambda_k,$$

and therefore

$$\partial G/\partial\lambda_k = \partial\lambda_1/\partial\lambda_k = (-1)^k J_k/J_1 \qquad \text{for } k \in [2, n].$$

If now, for some $k \in [2, n]$, $(-1)^k J_k/J_1 > 0$, then we could find $\mathbf{s} \in U$ such that

$$\lambda_i(\mathbf{s}) = \lambda_i(\mathbf{t}) \qquad \text{for } i \in [2, n]\backslash\{k\}$$

while

$$\lambda_i(\mathbf{s}) < \lambda_i(\mathbf{t}) \qquad \text{for both } i = 1 \text{ and } i = k.$$

Hence, for an optimal $\mathbf{t}$, $\mathbf{s}$ would also be optimal, yet $\Lambda_{\mathbf{s}}$ would not equioscillate, contradicting Kilgore's result. This proves (7) for an optimal $\mathbf{t}$ and so proves (8) and Lemma 3.

LEMMA 4. *The map Γ takes ∂T into $\partial \mathbb{R}^{n-1}$. Explicitly, if $\mathbf{t} \to \mathbf{s} \in \bar{T}$ with $\Delta s_i = 0$ for some $i \in [0, n-1]$, then $\|\Gamma(\mathbf{t})\| \to \infty$.*

Proof. Since $\sum \Delta s_j = b - a \neq 0$, there exists i such that $\Delta s_i = 0$ while either Δs_{i-1} or Δs_{i+1} is not zero. Assume without loss of generality that $\Delta s_i = 0$ and $\Delta s_{i-1} \neq 0$. Now pick $\hat{t} := (t_{i-1} + t_i)/2$ and let x be an arbitrary point in (t_i, t_{i+1}). Then

$$\left|\frac{\hat{t} - t_r}{x - t_r}\right| \geqslant \begin{cases} \dfrac{\hat{t} - t_{i-1}}{t_{i+1} - t_{i-1}} & \text{for} \quad r \leqslant i - 1 \\ \dfrac{1}{2}\dfrac{\Delta t_{i-1}}{\Delta t_i} & \text{for} \quad r = i, i+1 \\ 1 & \text{for} \quad r \geqslant i + 1. \end{cases}$$

Therefore, for all $j \in [0, n]$,

$$\left|\frac{l_j(\hat{t})}{l_j(x)}\right| = \prod_{r \neq j} \left|\frac{\hat{t} - t_r}{x - t_r}\right| \geqslant \frac{1}{2}\frac{\Delta t_{i-1}}{\Delta t_i} \prod_{r<i} \frac{\hat{t} - t_{i-1}}{t_{i+1} - t_{i-1}} \to \infty$$

as $\Delta t_i \to 0$ and $\Delta t_{i-1} \to \Delta s_{i-1} \neq 0$. This shows that

$$\lim_{\mathbf{t} \to \mathbf{s}} \Lambda_{\mathbf{t}}(\hat{t})/\Lambda_{\mathbf{t}}(x) = \infty \qquad \text{for every} \quad x \in (t_i, t_{i+1}).$$

Therefore $\lim_{\mathbf{t} \to \mathbf{s}} \lambda_i(\mathbf{t})/\lambda_{i+1}(\mathbf{t}) = \infty$, and so $\lim_{\mathbf{t} \to \mathbf{s}} (\lambda_{i+1} - \lambda_i)(\mathbf{t}) = -\infty$ since $\lambda_{i+1} \geqslant 1$. This proves that $\lim_{\mathbf{t} \to \mathbf{s}} \|\Gamma(\mathbf{t})\| = \infty$ and so proves the lemma.

Theorem 1 is an immediate consequence of Lemmas 3 and 4 and of the following.

THEOREM A (see, e.g., [2, 10]). *A local homeomorphism f of $\mathbb{R}^m$ to $\mathbb{R}^m$ with $\lim_{\|x\| \to \infty} \|f(x)\| = \infty$ is a homeomorphism of $\mathbb{R}^m$ onto $\mathbb{R}^m$.*

In a certain sense, this theorem is trivial since it is a special case of well-known facts regarding covering maps: The function f is a covering map for

$\mathbb{R}^m$ and so, since $\mathbb{R}^m$ is connected and simply connected, f is a universal covering map, therefore equivalent to any other universal covering map for $\mathbb{R}^m$, in particular, f is equivalent to the identity on $\mathbb{R}^m$ (see, e.g., [13, pp. 80–81]). But, for completeness, we now give an outline of a direct proof of the theorem.

The range of f is open, since f is locally 1–1 hence an open map. The range of f is also closed since $\lim f(x_r) = \alpha$ implies that the sequence $(f(x_r))$ is bounded, therefore, since f "maps ∞ to ∞" by assumption, (x_r) is bounded, hence can be assumed to converge to some x for which then $f(x) = \alpha$. This shows that the range of f is $\mathbb{R}^m$.

To show that f is 1–1, assume that $f(x) = f(y)$ for some $x, y \in \mathbb{R}^m$. The function $h : I \times I \to \mathbb{R}^m : (s, t) \mapsto (1 - t)\, h_0(s) + tf(x)$ with $h_0 : I \to \mathbb{R}^m : s \mapsto f(sx + (1 - s)\, y)$ and $I := [0, 1]$ is then a continuous map for which $h(z) = f(x)$ for all z in the set

$$B := (\{0\} \times I) \cup (I \times \{1\}) \cup (\{1\} \times I).$$

But now, the assumptions on f allow one to "lift" the map h, i.e., to show the existence of a continuous map $g : I \times I \to \mathbb{R}^m$ so that $f \circ g = h$ and $g(0, 0) = y$, therefore $g(s, 0) = sx + (1 - s)\, y$ for all $s \in I$. This implies that both x and y belong to the connected set $g(B)$ on which f is constantly equal to $f(x)$, and the fact that f is locally 1–1 now implies that $x = y$.

This proves the theorem, except for the technical part of "lifting" h. But this can be proved e.g., as is Lemma 3 of [13, p. 71] after one has proved, as in the proof of Theorem 2 below, that curves can be lifted uniquely.

We now prove Erdös's conjecture that, for every $\mathbf{t} \in T$,

$$\lambda^* \in [\min_i \lambda_i(\mathbf{t}), \max_i \lambda_i(\mathbf{t})].$$

THEOREM 2. *If* $\lambda_i(\mathbf{s}) \leqslant \lambda_i(\mathbf{t})$ *for* $i = 1,..., n$, *then* $\mathbf{s} = \mathbf{t}$.

Proof. If $\lambda_i(\mathbf{s}) = \lambda_i(\mathbf{t})$ for all i, then $\mathbf{s} = \mathbf{t}$ by Theorem 1. Hence assume that $\lambda_k(\mathbf{s}) < \lambda_k(\mathbf{t})$ for some k. This leads to a contradiction as follows.

The map $f : T \to \mathbb{R}^{n-1} : \mathbf{r} \to \boldsymbol{\lambda}(\mathbf{r}) := (\lambda_i(\mathbf{r}))_2^n$ is a local homeomorphism since $\det f'(\mathbf{r}) = J_1(\mathbf{r}) \neq 0$ for all $\mathbf{r} \in T$. We can therefore "lift" any continuous curve $h : [0, 1] \to \mathbb{R}^{n-1}$ to a curve in T as long as λ_1 stays bounded "along" h. Specifically, let

$$h : [0, 1] \to \mathbb{R}^{n-1} : \alpha \mapsto (1 - \alpha)\, \boldsymbol{\lambda}(\mathbf{s}) + \alpha \boldsymbol{\lambda}(\mathbf{t}).$$

Since f is locally 1–1, there exists, for each $\alpha \in [0, 1]$, at most one continuous function $g_\alpha : [0, \alpha] \to T$ so that $g_\alpha(0) = \mathbf{s}$ and $f \circ g_\alpha = h$ on $[0, \alpha]$. Let A be the set of such α. Then A is not empty since it contains 0. Further, A is open since, for every $\alpha \in [0, 1)$, some neighborhood V of $g_\alpha(\alpha)$ is mapped 1–1 onto a ball around $h(\alpha)$ by f, hence g_α can be extended continuously to the interval $[0, \alpha] \cup h^{-1} \circ f(V)$ which contains α in its interior. Finally, A is closed. To see this, it is sufficient to prove that $[0, \hat{\alpha}) \subseteq A$ implies $\hat{\alpha} \in A$, which can be done as follows. Since $[0, \hat{\alpha}) \subseteq A$, $g : [0, \hat{\alpha}) \to T : \alpha \mapsto g_\alpha(\alpha)$ defines a continuous map with $g(0) = \mathbf{s}$ and $f \circ g = h$ on $[0, \hat{\alpha})$. We claim that $g(\alpha)$ converges to some point in T as $\alpha \to \hat{\alpha}$. Indeed, for $i \in [2, n]$, $\lambda_i(g(\alpha))$ increases toward $h_i(\hat{\alpha}) = (1 - \hat{\alpha})\,\lambda_i(\mathbf{s}) + \hat{\alpha}\lambda_i(\mathbf{t})$ as $\alpha \to \hat{\alpha}$. Therefore, by (8) in the proof of Lemma 3, $\lambda_1(g(\alpha))$ decreases monotonely as $\alpha \to \hat{\alpha}$, hence it must have a limit since it is bounded below (by 1, for instance). This shows that $\lim_{\alpha \to \hat{\alpha}} \Gamma(g(\alpha))$ exists in $\mathbb{R}^{n-1}$, hence $g(\alpha)$ converges to some point $\mathbf{r} \in T$, by Theorem 1. But then, the definition $g(\hat{\alpha}) := \mathbf{r}$ provides a continuous extension of g to $[0, \hat{\alpha}]$ with $fg(\hat{\alpha}) = h(\hat{\alpha})$, hence $\hat{\alpha} \in A$.

This shows that $A = [0, 1]$, hence there exists $g : [0, 1] \to T$ continuous so that $g(0) = \mathbf{s}$ and $f \circ g = h$. Therefore, with $\mathbf{r} := g(1)$, we have $\lambda_i(\mathbf{r}) = \lambda_i(\mathbf{t})$ for all $i \in [2, n]$, while $\lambda_1(\mathbf{r}) \leqslant \lambda_1(\mathbf{s}) \leqslant \lambda_1(\mathbf{t})$. But, since $\lambda_k(\mathbf{s}) < \lambda_k(\mathbf{t})$ for some k, it follows that actually

$$\lambda_1(\mathbf{r}) < \lambda_1(\mathbf{t}),$$

either because $k = 1$, or else because λ_k strictly increases along the curve g, therefore λ_1 must strictly decrease along that curve, by (8) in the proof of Lemma 3.

Consider now the curve

$$h\colon [0, \infty) \to \mathbb{R}^{n-1}\colon \alpha \mapsto (\lambda_i(\mathbf{r}) - \alpha)_2^n.$$

By the preceding argument, there exists $\hat{\alpha} > 0$ and a continuous function $g : [0, \hat{\alpha}) \to T$ so that $f \circ g(\alpha) = (\lambda_i(\mathbf{r}) - \alpha)_2^n$ for all $\alpha < \hat{\alpha}$, while $\lambda_1(g(\alpha))$ strictly increases from $\lambda_1(\mathbf{r})$ at $\alpha = 0$ to ∞ at $\alpha = \hat{\alpha}$. This implies that

$$(\lambda_{i+1} - \lambda_i)(g(\alpha)) = (\lambda_{i+1} - \lambda_i)(\mathbf{r}) = (\lambda_{i+1} - \lambda_i)(\mathbf{t}) \qquad \text{for all } i \in [2, n-1]$$

while $(\lambda_2 - \lambda_1)(g(\alpha)) = \lambda_2(\mathbf{t}) - \alpha - \lambda_1(g(\alpha))$ decreases from its value $(\lambda_2(\mathbf{t}) - \lambda_1(\mathbf{r}))$ at $\alpha = 0$ to $-\infty$. But since $\lambda_1(\mathbf{r}) < \lambda_1(\mathbf{t})$, there exists therefore α so that $(\lambda_2 - \lambda_1)(g(\alpha)) = (\lambda_2 - \lambda_1)(\mathbf{t})$. But then $\Gamma(\mathbf{t}) = \Gamma(g(\alpha))$ while $g(\alpha) \neq \mathbf{t}$ since, e.g., $\lambda_2(g(\alpha)) < \lambda_2(\mathbf{t})$. This contradiction to Theorem 1 finishes the proof of Theorem 2.

COROLLARY. *For all $k \in [1, n]$, the map $\Gamma_k : T \to \mathbb{R}^{n-1} : \mathbf{r} \mapsto (\lambda_i(\mathbf{r}))_{i \neq k}$ is (globally) one–one.*

Proof. If $\Gamma_k(\mathbf{r}) = \Gamma_k(\mathbf{s})$, then either $\lambda_i(\mathbf{r}) \leqslant \lambda_i(\mathbf{s})$ for all i or else $\lambda_i(\mathbf{r}) \geqslant \lambda_i(\mathbf{s})$ for all i, hence $\mathbf{r} = \mathbf{s}$ by Theorem 2.

We note that Theorem 2 provides another proof of the characterization of the optimal node vector $\mathbf{t}$ as the unique point in T for which $\Lambda_\mathbf{t}$ equioscillates. Theorem 2 also shows that the optimal node vector is of no practical importance. For Brutman [3] has recently shown that, with

$$t_i = \left(a + b + (a - b)\left[\cos\frac{2i+1}{2n+2}\pi\right]\Big/\cos\frac{\pi}{2n+2}\right)\Big/2, \qquad i \in [0, n], \tag{9}$$

the zeros of the Chebyshev polynomial of degree $n + 1$, adjusted to the interval $[a, b]$ in such a way that the first and the last zero fall on the end points of the interval,

$$\max_i \lambda_i(\mathbf{t}) - \min_i \lambda_i(\mathbf{t}) \leqslant 0.201.$$

Numerical evidence strongly indicates that even

$$\max_i \lambda_i(\mathbf{t}) - \min_i \lambda_i(\mathbf{t}) < 0.0196$$

which would mean that the easily constructed node vector (9) produces an interpolation operator $P_{\mathbf{t}}$ whose norm is within 0.02 of the best possible value for all n.

4. Trigonometric Interpolation

In this section, we carry over the analysis of Sections 2 and 3 to the case of interpolation by trigonometric polynomials, i.e., by elements of

$$\mathbb{T}_n := \text{span}\{1, \cos x, \sin x, ..., \cos nx, \sin nx\},$$

on $[0, 2\pi)$. Because of the periodicity, the problem is altered slightly. Corresponding to each point $\mathbf{t}$ in

$$T := \{\mathbf{t} \in \mathbb{R}^{2n} : 0 < t_1 < t_2 < \cdots < t_{2n} < 2\pi\},$$

we construct the linear map $P_{\mathbf{t}}$ of trigonometric interpolation in $C[0, 2\pi)$ at the $2n + 1$ points $0 =: t_0 < \cdots < t_{2n} < 2\pi$. In its Lagrange form,

$$P_{\mathbf{t}} f = \sum_{i=0}^{2n} f(t_i)\, l_i$$

with

$$l_i(x) := \prod_{\substack{k=0\\k\neq i}}^{2n} \frac{S(x - t_k)}{S(t_i - t_k)}, \qquad \text{for all} \quad i \in [0, 2n].$$

Here, we use the abbreviation

$$S(x) := \sin(x/2).$$

We have again $\| P_{\mathbf{t}} \| = \| \Lambda_{\mathbf{t}} \|$ where $\Lambda_{\mathbf{t}} := \sum_i | l_i |$. Set

$$\lambda_i(\mathbf{t}) := \max_{t_{i-1} \leqslant x \leqslant t_i} \Lambda_{\mathbf{t}}(x), \qquad \text{for all} \quad i \in [1, 2n+1],$$

with $t_{2n+1} := 2\pi$.

THEOREM 3. *We have* $\| P_{\mathbf{t}} \| = \lambda^* := \inf_{\mathbf{s} \in T} \| P_{\mathbf{s}} \|$ *exactly when* $\mathbf{t} = \mathbf{t}^* := (i/(2n+1))_1^{2n}$, *in which case* $\Lambda_{\mathbf{t}}$ *equioscillates. Furthermore, for any* $\mathbf{t} \in T \setminus \{\mathbf{t}^*\}$,

$$\min_i \lambda_i(\mathbf{t}) < \lambda^* < \max_i \lambda_i(\mathbf{t}).$$

Proof. We begin with a proof of the claim that

$$\det(\partial\lambda_i(\mathbf{t})/\partial t_j)_{i=1;\, i \neq k}^{2n+1}{}_{j=1}^{2n} \neq 0 \qquad \textit{for all} \quad \mathbf{t} \in T, \quad k \in [1, 2n+1]. \tag{10}$$

Let F_i be the unique trigonometric polynomial of degree n which agrees with $\Lambda_{\mathbf{t}}$ on $[t_{i-1}, t_i]$, for $i \in [1, 2n+1]$. Thus,

$$F_i(t_j) = \begin{cases} (-1)^{i-1-j} & \text{for } j \in [0, 1-1] \\ (-1)^{j-i} & \text{for } j \in [i, 2n+1]. \end{cases}$$

Let τ_i denote the unique point in $[t_{i-1}, t_i]$ at which $\Lambda_{\mathbf{t}}$, and hence F_i, takes on the value $\lambda_i(\mathbf{t})$. Now

$$\partial\lambda_i/\partial t_j = -F_i'(t_j)\, l_j(\tau_i) = \prod_{k=0}^{2n} S(\tau_i - t_k) \frac{F_i'(t_j)}{S(t_j - \tau_i)} \Big/ \prod_{\substack{k=0 \\ k \neq j}}^{2n} S(t_j - t_k)$$

which shows that λ_i is a continuously differentiable function on T and also shows that (10) is equivalent to

$$\det(q_i(t_j))_{i=1;\, i \neq k}^{2n+1}{}_{j=1}^{2n} \neq 0 \qquad \textit{for all} \quad \mathbf{t} \in T, \quad k \in [1, 2n+1], \tag{11}$$

where

$$q_i(x) := F_i'(x)/S(x - \tau_i), \qquad i \in [1, 2n+1].$$

For the proof of (11), we make use of the following result corresponding to Lemma 6 of [8]. Denote by $\tau_1^{(i)}, \ldots, \tau_{2n}^{(i)}$ the zeros of F_i' in $[0, 2\pi)$, necessarily all simple, in order.

LEMMA 5. *The zeros of* $F_1', \ldots, F_{2n+1}'$ *lie in the pattern*

$$\begin{aligned} 0 \leqslant \tau_{2n}^{(i)} < \tau_{2n}^{(i-1)} < \cdots < \tau_{2n}^{(1)} < \tau_1^{(2n+1)} < \cdots \\ < \tau_{2n-1}^{(1)} < \tau_{2n}^{(2n+1)} < \cdots < \tau_{2n}^{(i+1)} < 2\pi \end{aligned}$$

for a certain $i \in [1, 2n]$. *Note that* $\tau_{2n}^{(1)} = \tau_1$, *and* $\tau_{k-1}^{(k)} = \tau_k$ *for* $k \in [2, 2n+1]$.

The proof of Lemma 5 follows exactly the same lines as the one given in Section 2 for Lemma 6 of [8] (including the use of the trigonometric analog of Markov's result), except that matters are a little easier since both F_i and F_i' have exactly $2n$ zeros in $[0, 2\pi)$, for all i.

In order to use Lemma 5 in a proof of (11) much as Kilgore used Lemma 6 of [8] in his proof of (4), we must first show that

$$0 \leqslant s_1 < \cdots < s_{2n} < 2\pi \quad \textit{and} \quad \sum_{i=1}^{2n+1} a_i q_i(s_j) = 0 \quad \textit{for all} \quad j \in [0, 2n]$$
$$\textit{implies that} \quad \sum_{i=1}^{2n+1} a_i q_i = 0. \tag{12}$$

For this, observe that $F_i'(x) = \text{const} \prod_{k=1}^{2n} S(x - \tau_k^{(i)})$, therefore

$$q_i(x) = \text{const} \prod_{\substack{k=1 \\ k \neq i-1}}^{2n} S(x - \tau_k^{(i)}) \qquad \text{for all} \quad i \in [1, 2n+1].$$

Here, $k \neq i - 1$ is meant to read $k \neq 2n$ in case $i = 1$. This shows that q_i is not 2π-periodic, but 4π-periodic, and odd about 2π, i.e., $q_i(x + 2\pi) = -q_i(x)$, all x. Furthermore, the function $p_i(x) := q_i(2x)$, all x, is in

$$\mathbb{T}_{2n-1} = \text{span}\{1, \cos x, \sin x, \ldots, \cos(2n-1)\, x, \sin(2n-1)\, x\}.$$

Therefore, the hypotheses of (12) imply that the element $\sum_i a_i p_i$ of $\mathbb{T}_{2n-1}$ vanishes at the $4n$ distinct points $\hat{s}_1, \ldots, \hat{s}_{4n}$ with

$$\hat{s}_j := \begin{cases} s_j/2 & \text{for } j \in [1, 2n] \\ s_j/2 + \pi & \text{for } j \in [2n+1, 4n], \end{cases}$$

and so $\sum_i a_i p_i = 0$, proving (12).

The proof of (11) proceeds now as the proof of (4) in Section 2, and, with (10) thus established, the reasoning in the proofs of Theorems 1 and 2 in Section 3 applies directly to finish the proof of Theorem 3.

We note in passing that Ehlich and Zeller [5] have proved a formula for λ^* in the trigonometric case,

$$\lambda^* = \left\{1 + 2 \sum_{k=0}^{n-1} \left(\sin \frac{(2k+1)\,\pi}{(2n+1)\,2}\right)^{-1}\right\} \Big/ (2n+1). \tag{13}$$

Finally, the above analysis applies without essential change to the case when we also fix t_{2n} at some point $b < 2\pi$ and consider the optimal choice of $t_1 < \cdots < t_{2n-1}$ in $(0, b)$ for trigonometric interpolation.

Acknowledgment

We are grateful to M. S. Crandall for pointing out to us Theorem A and for joining us in the construction of a proof.

Added in proof. After completion of this work in March 1977, we received word from Theodore Kilgore that he, too, had succeeded in proving Bernstein's conjecture. His argument is given in [14] and proceeds along different lines.

References

1. S. N. Bernstein, Sur la limitation des valeurs d'une polynome $P(x)$ de degré n sur tout un segment par ses valeurs en $(n+1)$ points du segment, *Izv. Akad. Nauk SSSR* **7** (1931), 1025–1050.
2. F. E. Browder, Covering spaces, fibre spaces, and local homeomorphisms. *Duke Math. J.* **21** (1954), 329–336.
3. L. Brutman, On the Lebesgue function for polynomial interpolation, *SIAM J. Numer. Anal.* **15** (1978), 694–704.
4. E. W. Cheney and K. H. Price, Minimal projections, *in* "Approximation Theory" (A. Talbot, Ed.), pp. 261–289, Academic Press, London, 1970.
5. H. Ehlich and K. Zeller, Auswertung der Normen von Interpolationsoperatoren, *Math. Ann.*, **164** (1966), 105–112.
6. P. Erdös, Some remarks on polynomials, *Bull. Amer. Math. Soc.* **53** (1947), 1169–1176.
7. P. Erdös, Problems and results on the theory of interpolation, I, *Acta Math. Acad. Sci. Hungar.* **9** (1958), 381–388.
8. T. A. Kilgore, Optimization of the norm of the Lagrange interpolation operator, *Bull. Amer. Math. Soc.* **83** (1977), 1069–1071.
9. T. A. Kilgore and E. W. Cheney, A theorem on interpolation in Haar subspaces, *Aequationes Math.* **14** (1976), 391–400.
10. A. Lelek and Jan Mycielski, Some condition for a mapping to be a covering, *Fund. Math.* **49** (1960/61), 295–300.
11. F. W. Luttmann and T. J. Rivlin, Some numerical experiments in the theory of polynomial interpolation, *IBM J. Res. Develop.* **9** (1965), 187–191.
12. V. A. Markov, Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen, *Math. Ann.* **77** (1916), 213–258.
13. E. H. Spanier, "Algebraic Topology," McGraw–Hill, New York, 1966.
14. T. A. Kilgore, A characterization of Lagrange interpolating projections with minimal Tchebycheff norm, *J. Approximation Theory* **24** (1978), 273–288.