Rev. Mat. Iberoam. 40 (2024), no. 6, 2251–2290
DOI 10.4171/RMI/1470

Approximation by polynomials with only real critical points

David L. Bishop

Abstract. We strengthen the Weierstrass approximation theorem by proving that any real-valued continuous function on an interval $I \subset \mathbb{R}$ can be uniformly approximated by a real-valued polynomial whose only (possibly complex) critical points are contained in I. The proof uses a perturbed version of the Chebyshev polynomials and an application of the Brouwer fixed point theorem.

1. Introduction

The Weierstrass approximation theorem [21] states that for any real-valued, continuous function f on a compact interval $I \subset \mathbb{R}$, and for any $\varepsilon > 0$, there is a real polynomial p so that $\|f - p\|_I = \sup_{x \in I} |f(x) - p(x)| < \varepsilon$. The statement does not say much about the behavior of p off the interval I, but for some applications of Weierstrass' theorem, it would be advantageous to know the location of all the critical points of p; for example, in polynomial dynamics, the behavior of the iterates of p depends crucially on the orbits of all the complex critical points of p. In [2] it is shown that one can restrict all the critical points (real and complex) to a thin rectangle $I \times [-\varepsilon, \varepsilon]$. In this paper, we prove that we can actually take $\varepsilon = 0$. The following is our main result.

Theorem 1.1 (Critically constrained Weierstrass theorem). *Suppose $f : I \to \mathbb{R}$ is a continuous function on a compact interval $I \subset \mathbb{R}$. Then for any $\varepsilon > 0$, there is a real polynomial p so that $\|f - p\|_I < \varepsilon$ and* $\mathrm{CP}(p) := \{z \in \mathbb{C} : p'(z) = 0\} \subset I$*, i.e., every real or complex critical point of p is inside I. If f is A-Lipschitz, then p may be taken to be CA-Lipschitz for some $C < \infty$ independent of f.*

Using dilation and translation, it is enough to prove Theorem 1.1 for the particular interval $I = [-1, 1]$, and this is the only case we will consider from this point on. Recall that f is A-Lipschitz on I if $|f(x) - f(y)| \le A|x - y|$ for all $x, y \in I$. For a Lipschitz function f, the derivative f' exists and satisfies $|f'| \le A$ almost everywhere, and $f(x) = f(a) + \int_a^x f'(t)\, dt$ (e.g., see Section 3.5 of [10]). By the usual Weierstrass theorem, polynomials are dense in $C_{\mathbb{R}}(I)$ (the space of continuous, real-valued functions on I), and

Mathematics Subject Classification 2020: 41A10 (primary); 30E10 (secondary).
Keywords: Weierstrass approximation, critical points, Chebyshev polynomials, uniform approximation, weak star convergence.

every polynomial is Lipschitz when restricted to a compact interval, so it suffices to prove Theorem 1.1 when f is Lipschitz.

The assumption that f is real valued is necessary; a similar result does not hold for complex valued functions. Eremenko and Gabrielov [9] proved that any complex valued polynomial with only real critical points is essentially real-valued itself. More precisely, they proved that any such polynomial p is of the form $p(z) = aq(z) + b$, where $q(z)$ is a real polynomial and $a, b \in \mathbb{C}$. (Their result characterizes rational functions with only real critical points, but specializes to polynomials as above.) Polynomials with only real critical values have played a role in various problems, e.g., density of hyperbolicity in dynamics [14], rigidity of conjugate polynomials [7], Smale's conjecture on solving polynomial systems [12], and Sendov's conjecture on the locations of the critical points of a polynomial in terms of its roots [4].

We also note that Theorem 1.1 is non-linear in nature. For example, x^3 and $(x-1)^3$ each have a single critical point, and these are both real. However, it is easy to check that the sum $x^3 + (x-1)^3$ has two complex critical points. So the set of real polynomials with all critical points in $I = [0, 1]$ is not a linear subspace of $C_{\mathbb{R}}(I)$. Thus many usual methods, such as duality or reducing to approximating a spanning set, do not apply. Moreover, Theorem 1.1 need not be true for general compact subsets of $\mathbb{R}$. See Section 13 for some disconnected sets where it fails.

If $p_n \to f$ uniformly, we might hope that $p_n' \to f'$, at least when f is analytic, but this is false, except in very special cases. To see why, we first recall that the Laguerre–Pólya class is the collection of entire functions (holomorphic functions on $\mathbb{C}$) that are limits, uniformly on compact sets, of real polynomials with only real zeros. These have been characterized as follows [17]: it is the collection of entire functions f so that (1) all roots are real, (2) the nonzero roots satisfy $\sum_n |z_n|^{-2} < \infty$, and (3) we have a Hadamard factorization

$$f(z) = z^m e^{a+bz+cz^2} \prod_n \left(1 - \frac{z}{z_n}\right) e^{z/z_n}, \tag{1.1}$$

with $m \in \{0, 1, 2, \ldots\}$, $a, b \in \mathbb{R}$ and $c \leq 0$. In particular, functions like $\exp(-z^2)$ and $\sin(z)$ are in the Laguerre–Pólya class, but $\exp(z^2)$ and $\sinh(z)$ are not. This class arises in many contexts, e.g., the Riemann hypothesis is equivalent to the claim that a certain explicit formula defines a function in the Laguerre–Pólya class [6, 19].

A theorem of Korevaar and Loewner [13], extending earlier work of Laguerre and Pólya, says that if $\{p_n\}$ are polynomials with only real zeros that converge uniformly to f on an interval $I \subset \mathbb{R}$, then f must be the restriction to I of a Laguerre–Pólya entire function, and that p_n converges to f on the whole complex plane (uniformly on compact sets). Clunie and Kuijlaars later proved that this also holds if we only assume p_n converges in measure to f on a subset $E \subset \mathbb{R}$ of positive measure (see Corollary 1.3 in [5]). Recall that $p_n \to f$ in measure if, for every $\varepsilon > 0$, $|\{x : |f(x) - p_n(x)| > \varepsilon\}|$ tends to zero (in this paper, $|E|$ will denote the Lebesgue measure of a measurable subset of $\mathbb{R}$.) For bounded intervals $I \subset \mathbb{R}$, pointwise convergence almost everywhere implies convergence in measure, so the same conclusion holds if $p_n \to f$ pointwise on a set E of positive measure. Thus there exist analytic functions f on I so that f' cannot be a limit of polynomials with only real zeros, either uniformly, in measure, or pointwise on a set of positive measure. However, our proof of Theorem 1.1 will also give the following result.

Theorem 1.2. *There is a $C < \infty$ so that every bounded, measurable function f on I is the weak-$*$ limit in L^∞ of a sequence of polynomials $\{p_n\}$ with only real zeros, and such that* $\sup_n \|p_n\|_I \leq C\|f\|_\infty$. *This fails for $C = 1$.*

Here $\|g\|_\infty$ denotes the L^∞ norm on I, and $p_n \to f$ weak-$*$ in $L^\infty(I, dx)$ if

$$\int p_n g\, dx \to \int fg\, dx$$

for every Lebesgue integrable function g on I.

The polynomials constructed in our proof of Theorem 1.2 will diverge pointwise almost everywhere, but this is not an artifact of the proof; it is forced in many cases. We claim that if f in Theorem 1.2 is not in the Laguerre–Pólya class, then $\{p_n(x)\}$ diverges almost everywhere on the set where f is non-zero. To prove this, suppose f is not in the Laguerre–Pólya class and that $\{p_n\}$ is uniformly bounded, has only real zeros, and converges weak-$*$ to f. The Clunie–Kuijlaars theorem stated above implies that if $\{p_n\}$ has pointwise limits on a set of positive measure, these limits must be zero almost everywhere. However, if a uniformly bounded sequence $\{p_n\}$ converges pointwise to 0 on a set E, then the dominated convergence theorem (see, e.g., Theorem 2.24 of [10]) implies that $\int_E p_n \to 0$. By weak-$*$ convergence of p_n to f, we then have

$$0 = \lim_n \int_E p_n = \lim_n \int p_n \chi_E = \int f\chi_E = \int_E f,$$

where we have taken $g = \chi_E$ in the definition of weak-$*$ convergence. As usual, χ is the characteristic (or indicator) function of E ($\chi = 1$ on E and $\chi = 0$ off E). Since this also holds for every measurable subset of E, we deduce f is zero almost everywhere on E (otherwise either $\int_{E\cap\{f>0\}} f$ or $\int_{E\cap\{f<0\}} f$ would be non-zero). This proves the claim.

In order to prove Theorem 1.1, we want to write f as the uniform limit of polynomials p_n whose derivatives have the form

$$p_n'(x) = C_n \prod_{k=1}^{n} (x - z_k^n), \tag{1.2}$$

where $C_n \in \mathbb{R}$ and $\{z_k^n\}_{k=1}^n \subset [-1, 1]$. These points will be perturbations of the roots of T_n, the degree n Chebyshev polynomial (of the first kind). We briefly recall the definition.

Let $J(z) = \frac{1}{2}(z + 1/z)$ be the Joukowsky map. Note that a point $z = x + iy$ on the unit circle is mapped to $x \in [-1, 1]$, and J is a 1-1 holomorphic map of $\mathbb{D}^* = \{z : |z| > 1\}$ to $U = \mathbb{C} \setminus [-1, 1]$. Thus it has a holomorphic inverse $J^{-1} : U \to \mathbb{D}^*$. Then $T_n = J((J^{-1})^n)$ is a n-to-1 holomorphic map of U to U that is continuous across $\partial U = [-1, 1]$. By Morera's theorem (see, e.g., Theorem 4.19 of [15]), such a function is entire (holomorphic on the whole plane). Since T_n is finite-to-1, Picard's great theorem (see, e.g., Theorem 10.14 of [15]) implies it is a polynomial, and since T_n is n-to-1, the fundamental theorem of algebra implies it must have degree n. Unwinding the definitions, T_n maps $[-1, 1]$ into itself, takes the extreme values ± 1 at the points $\{x_k^n\} = \{\cos(\pi \frac{k}{n})\}_{k=0}^n$ (the vertical projections of the nth roots of unity), and it has its roots at $\{r_n^k\} = \{\cos(\pi \frac{2k-1}{2n})\}_{k=1}^n$ (the vertical projections of the midpoints between the roots of unity). More background and facts about the Chebyshev polynomials will be given in Section 2.

The basic idea of the proof of Theorem 1.1 is to consider polynomials as in (1.2) where $z_k^n = r_k^n + y_k^n$ are small perturbations of the Chebyshev roots. Fix a large positive integer n and consider the Chebyshev polynomial T_n. For $k = 1, \dots, n-1$, let $I_k^n = [r_k^n, r_{k+1}^n]$ denote the interval between the kth and $(k+1)$st roots of T_n. We call these "nodal intervals" and call the part of the graph of T_n above I_k^n a "node" of T_n. Every node of T_n is either positive or negative. Suppose it is positive. If we move the roots at the endpoints of I_k^n farther apart (but leave all the other roots of the Chebyshev polynomial unchanged), then the node between them becomes higher, and the two adjacent negative nodes each becomes smaller (less negative). Thus the integral of the new polynomial over the union of these three intervals becomes more positive. See Figure 1. This figure, and many others in this paper, was drawn using the MATLAB program Chebfun by L. N. Trefethen and his collaborators. See [11].

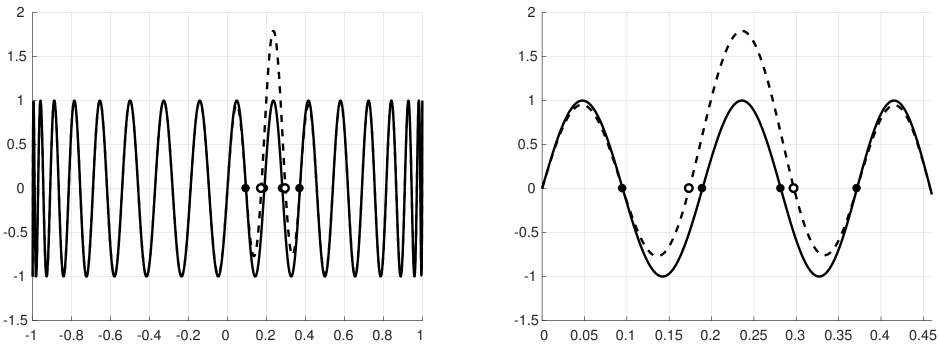

Figure 1. A 2-point perturbation of T_{33}. The left picture shows all of $[-1, 1]$ and the right shows an enlargement of the interval where the perturbation occurs. The Chebyshev polynomial is solid and the perturbation is dashed. The white dots are the two new root locations.

When we move each endpoint of I_k^n by $t|I_k^n|$, the integral of the polynomial over I_k^n changes by at least some fixed multiple of $t|I_k^n|$. A quantitative estimate like this is one of the key results of this paper, although we shall give it for perturbations involving three roots instead of only two. We prefer the (more complicated) 3-point perturbations, because we can choose them so that the effect on T_n far from the perturbed roots decreases more quickly (like d^{-3} instead of d^{-2}, where d is the distance to the perturbed roots). A precise estimate is formulated and proven in Section 6. See Figure 2 for an example of a 3-point perturbation.

See Figure 3 for a degree 33 approximation to $f(x) = |x|$. A degree 201 approximation is shown in Figure 4, which also shows a log-log plot showing the rate of approximation versus the degree of the polynomial. Our approximations were chosen by enlarging negative nodes to the left of the origin and enlarging positive nodes to the right, but no attempt was made to do this in an optimal way. Nevertheless, the rate of approximation is approximately the reciprocal of the degree. This is a little surprising, since the best sup-norm approximation of $f(x) = |x|$ by a degree n polynomial (with no restrictions on the critical points) satisfies $\|f - p_n\| \sim (.280169)/n$, e.g., see Chapter 25 of [20]. Thus

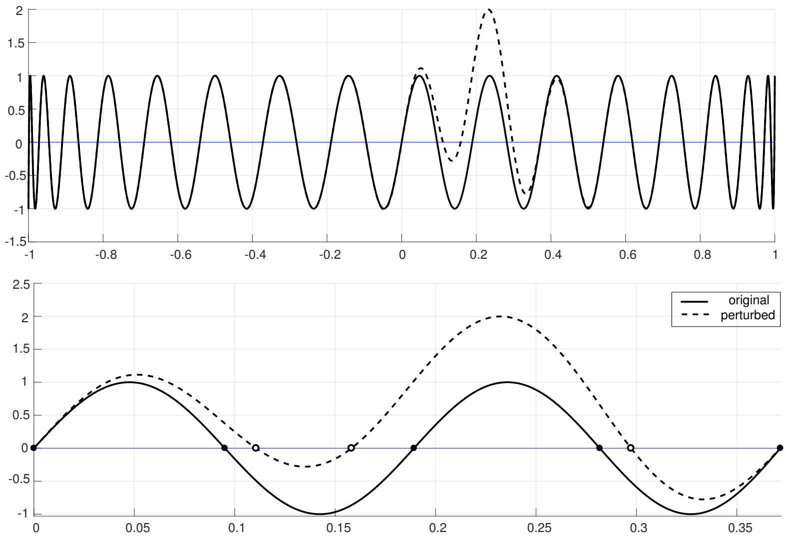

Figure 2. A 3-point perturbation. The top figure shows T_{33} (solid) and the perturbation $\widetilde{T}_n$ (dashed) on $[-1, 1]$. The bottom figure is an enlargement around the perturbed roots.

our approximations (which are just a first guess) are fairly close to the best approximation. Figure 5 gives another example of approximating a Lipschitz function by weakly approximating its derivative.

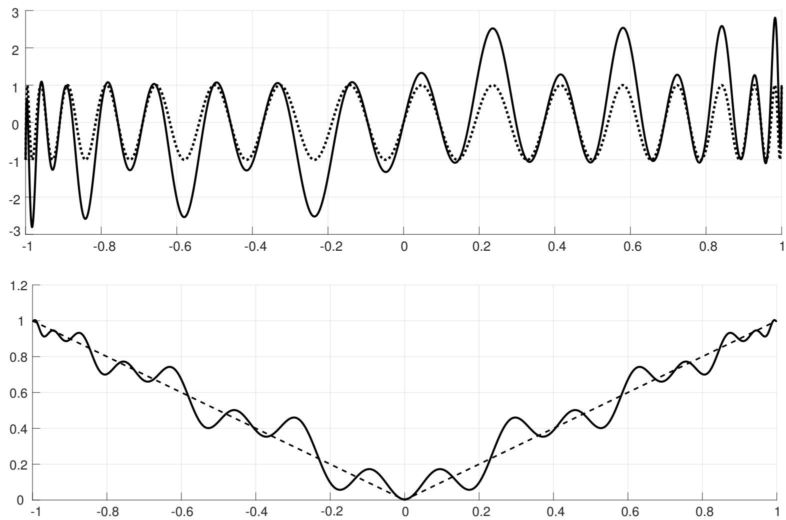

Figure 3. On top we have perturbed T_{33} (dashed) to obtain p' (solid): four pairs in $[0, 1]$ chosen to make the function more positive, and four pairs in $[-1, 0]$ chosen to make it more negative. The bottom picture shows $p = \int p'$ (solid), which approximates $f(x) = |x|$ (dashed). See Figure 4 for a higher degree approximation.

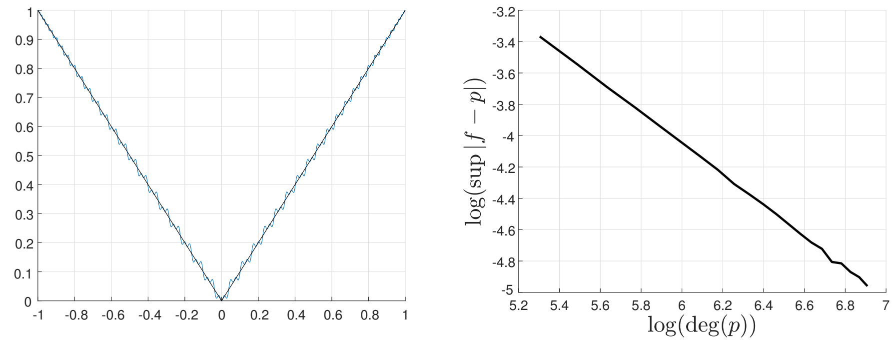

Figure 4. On the left is a degree 201 polynomial approximating $|x|$. The right picture is a log-log plot of the sup-norm difference between $f(x) = |x|$ and our approximation for degrees between 200 and 1000. The best linear fit is $\approx (-.9912)t + 1.8972$, where $t = \log(\deg(p))$. The optimal polynomial approximations (no restrictions) behave like $\approx -t - 1.2724$.

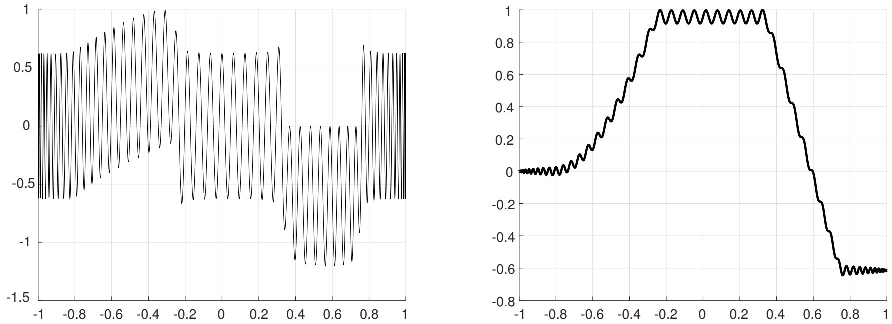

Figure 5. On the left is a perturbed Chebyshev polynomial of degree 100, and on the right is its integral. From the picture it seems clear that any Lipschitz function can be approximated; the goal of the paper is to prove this is correct.

Briefly, the proof of Theorem 1.1 will proceed as follows. We convert the $n-1$ nodal intervals into $N = (n-1)/4$ larger intervals, $\{G_k^n\}_{k=1}^N$, by taking unions of groups of four adjacent nodal intervals. We would like the origin to be the common endpoint of two such intervals, and the whole arrangement to be symmetric with respect to the origin, and this leads us to assume $n-1$ is a multiple of eight. We then estimate how the Chebyshev polynomial changes when we slightly perturb the three interior roots in a single interval G_k^n. We make precise the idea that the change is large inside G_k^n, and small outside this interval (and decays as we move away from G_k^n). Most of the computations are done when G_k^n is linearly rescaled to be approximately $[-2, 2]$, but these estimates are easily converted to estimates on the original intervals. These estimates will show that there is a $t > 0$ so that, for any vector $y = (y_1, \ldots, y_N)$ with coordinates $|y_k| \le t$, there is a perturbation of the roots of T_n that lie in the interior of G_k^n so that the integral of the perturbed polynomial over G_k^n equals $y_k \cdot |G_k^n|$. Of course, perturbations of roots in other intervals may destroy this equality, but using the Brouwer fixed point theorem, we will show that there is a perturbation of all the roots that gives the desired equality over every G_k^n simultaneously (except for a bounded number of exceptions near ± 1).

In order to prove Theorem 1.1, it suffices to consider functions f with a small Lipschitz constant, e.g., less than the value t chosen above. For each interval G_k^n, we take $y_k = \Delta(f, G_k^n)/|G_k^n|$, where $\Delta(f, [a, b]) = f(b) - f(a)$. Then $|y_k| \le t$ for all k, since f is t-Lipschitz. Using Brouwer's theorem, we can therefore perturb the roots of T_n to obtain a perturbed polynomial $T_n(x, y)$ so that $\int_{G_k^n} T_n(x, y)dx = y_k|G_k^n|$ for every k. Then any anti-derivative F of the perturbed polynomial satisfies $\Delta(F, G_k^n) = \Delta(f, G_k^n)$ for every k, so choosing an anti-derivative such that $F(0) = f(0)$, implies that F equals f at every endpoint of every G_k^n (again, with a small number of exceptions near ± 1). Since both F and f are Lipschitz with bounds independent of n, and since $|G_k^n| \to 0$ as $n \nearrow \infty$, this implies F uniformly approximates f when n is large enough, proving Theorem 1.1.

Roughly speaking, the remainder of the paper divides into four parts. Part I: Sections 2–3 describe basic properties of Chebyshev polynomials and their nodal intervals. Part II: Sections 4-7 define the perturbations $T_n(x, y)$ of the Chebyshev polynomials T_n and give estimates for how the perturbed polynomials compare to $T_n(x)$. Part III: Sections 8-9 verify the conditions needed to apply Brouwer's theorem and we prove Theorem 1.1 in Section 10. Part IV gives some auxiliary results: Section 11 proves Theorem 1.2, Section 12 shows that our polynomial approximants have derivatives that diverge almost everywhere, and an example of how Theorem 1.1 can fail for some disconnected sets is given in Section 13.

When A and B are both quantities that depend on a common parameter, then we use the usual notation $A = O(B)$ to mean that the ratio B/A is bounded independently of the parameter. The more precise notation $A = O_C(B)$ will mean $|A| \le C|B|$. For example, $x = 1 + O_2(1/n)$ is simply a more concise way of writing $1 - 2/n \le x \le 1 + 2/n$. The notation $A = \Omega_C(B)$ means $A \ge C|B|$ or, equivalently, $B = O_C(A)$. We write $A \simeq B$ if both $A = O(B)$ and $B = O(A)$.

2. Estimating the length of the nodal intervals

The Chebyshev polynomials defined in the introduction have a number of alternate definitions, see e.g. [20]. For $|x| \le 1$, we can write

$$\begin{aligned} T_n(x) = \cos(n \arccos(x)) &= \frac{1}{2}\Big[\big(x - \sqrt{x^2 - 1}\big)^n + \big(x + \sqrt{x^2 - 1}\big)^n\Big] \\ &= \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} (x^2 - 1)^k x^{n-2k}, \end{aligned}$$

or $\{T_n\}$ can be defined by the three term recurrence

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{n+1}(x) = 2x\,T_n(x) - T_{n-1}(x) \quad \text{for } n \ge 1.$$

The latter makes it clear that

$$T_n(x) = 2^{n-1} \prod_k (x - r_k^n),$$

where $\{r_k^n\}$ are the Chebyshev roots defined in the introduction. In particular, the coefficient of x^n in T_n is 2^{n-1}. Among polynomials of degree n with leading coefficient 2^{n-1}, T_n minimizes the supremum norm over $[-1, 1]$; this is the min-max property (a special case is proven in Lemma 7.1). Also note that the sup of $|T_n|$ over $[-1, 1]$ is 1. As mentioned in the introduction, T_n has $n-1$ critical points, all with singular values -1 or 1, and the endpoints are also extreme points with $T_n(1) = 1$, $T_n(-1) = (-1)^n$. When we perturb the roots of T_n, some of these extremal values must increase in absolute value, and the construction in this paper is based on controlling where and how much this happens.

The Chebyshev polynomials are orthogonal with respect to $d\mu = dx/\sqrt{1-x^2}$, and expansions in terms of Chebyshev polynomials are extremely useful in numerical analysis; indeed, Chebyshev polynomials are the direct analog on $[-1, 1]$ of Fourier expansions on the circle, and many theorems about Fourier expansions transfer to Chebyshev expansions. Chebyshev polynomials are also well behaved under multiplication and composition, i.e.,

$$T_n(x)T_m(x) = \frac{1}{2}\left[T_{m+n}(x) + T_{|m-n|}(x)\right] \quad \text{and} \quad T_n(T_m(x)) = T_{mn}(x).$$

With the additional condition $\deg(T_n) = n$, the latter characterizes (up to a linear change of variable) Chebyshev polynomials and the power functions $\{x^n\}$, see [18].

In this paper, we will not need most of the properties above, but we will need precise estimates of the lengths of the nodal intervals $\{I_k^n\}$, and of the integral of T_n over these nodal intervals. The length estimates are addressed in this section, and the area estimates in the following section.

Let

$$J_k^n = \left[\pi\,\frac{2k-1}{2n}, \pi\,\frac{2k+1}{2n}\right], \quad \text{for } k = 1, \dots, n-1.$$

Each of these intervals has length π/n. Let $I_k^n = -\cos(J_k^n)$, $k = 1, \dots, n-1$, be the nodal interval between the kth and $(k+1)$st roots of T_n, i.e., $I_k^n = [r_k^n, r_{k+1}^n]$. (We introduce the minus sign so the I_k^n are labeled left to right in $[-1, 1]$.) Note that when n is even, there are an odd number of nodal intervals and $I_{n/2}^n$ contains the origin as its midpoint. When n is odd, the intervals $I_{(n-1)/2}^n$ and $I_{(n+1)/2}^n$ share the origin as an endpoint. In both cases, the intervals with $k \le n/2$ cover $[-1, 0]$. In our application, we always take n odd, but the estimates in the section apply to both cases.

Let $|I|$ denote the length of an interval I. Our first goal is to establish some basic facts about lengths of the nodal intervals and the distances between them. Note that, by symmetry, $|I_k^n| = |I_{n-k}^n|$, so most of our estimates are only given for $1 \le k \le (n-1)/2$, i.e., subintervals of $[-1, 0]$.

Lemma 2.1. $|I_k^n| \le |J_k^n| = \pi/n$.

Proof. Clearly I_k^n is the vertical projection of $\exp(iJ_{n-k}^n)$, which has arclength π/n. ∎

The following says the biggest intervals are adjacent to the origin, and that the lengths monotonically decrease as we move out towards the endpoints.

Lemma 2.2. *For $x \in [r_1^n, r_{n-1}^n]$, let I_x be the nodal interval containing x (if x is the common endpoint of two nodal intervals, then take I_x to be the nodal interval containing x and closer to* 0*). Then $|x| < |y|$ implies $|I_x| \ge |I_y|$.*

Proof. This is also obvious since the intervals in question are vertical projections of equal length arcs on the unit circle, and the slope of the circle increases as we move toward either ± 1 from 0. ∎

Lemma 2.3. *For* $1 \le k \le (n-1)/2$, $4k/n^2 \le |I_k^n| \le k\pi^2/n^2 \approx (9.8696)\,k/n^2$.

Proof. Note that I_k^n has left endpoint $\cos(\pi\frac{2k+1}{2n})$ and right endpoint $\cos(\pi\frac{2k-1}{2n})$. From the difference rule for cosine,

$$\cos\alpha - \cos\beta = -2\sin\left(\frac{\alpha+\beta}{2}\right)\sin\left(\frac{\alpha-\beta}{2}\right), \tag{2.1}$$

we can deduce

$$\cos\left(\pi\frac{2k-1}{2n}\right) - \cos\left(\pi\frac{2k+1}{2n}\right) = 2\sin\left(\pi\frac{2k}{2n}\right)\cdot\sin\left(\frac{\pi}{2n}\right).$$

Now use $2x/\pi \le \sin x \le x$ on $[0,\pi/2]$ to derive the estimate in the lemma. ∎

Lemma 2.4. *With notation as above, if* $1 \le k \le k+j \le n/2$, *then*

$$1 \le \frac{|I_{k+j}^n|}{|I_k^n|} \le 1 + \frac{\pi}{2}\frac{j}{k}.$$

Proof. The left-hand inequality is just Lemma 2.2. Recall $J_k^n = [a,b] = [\pi\frac{2k-1}{2n}, \pi\frac{2k+1}{2n}]$ and $J_{k+j}^n = [c,d] = [\pi\frac{2k+2j-1}{2n}, \pi\frac{2k+2j+1}{2n}]$. Using (2.1), we get

$$\frac{|I_{k+j}|}{|I_k|} = \frac{\cos c - \cos d}{\cos a - \cos b} = \frac{\sin\left(\frac{c+d}{2}\right)\sin\left(\frac{c-d}{2}\right)}{\sin\left(\frac{a+b}{2}\right)\sin\left(\frac{a-b}{2}\right)} = \frac{\sin\left(\frac{c+d}{2}\right)}{\sin\left(\frac{a+b}{2}\right)},$$

since $(d-c)/2 = (b-a)/2$. Thus since $k/n \le 1/2$, $(\sin x)' = \cos x \le 1$ and $\sin x \ge 2x/\pi$ on $[0,\pi/2]$, we have

$$\begin{aligned}\frac{|I_{k+j}|}{|I_k|} &\le \frac{\sin(\pi(k+j)/n)}{\sin(\pi k/n)} = \frac{\sin(\pi k/n) + \pi j/n}{\sin(\pi k/n)} \\ &= 1 + \frac{\pi j/n}{\sin(\pi k/n)} \le 1 + \frac{\pi j/n}{(2/\pi)\pi k/n} = 1 + \frac{\pi}{2}\frac{j}{k}.\end{aligned}$$ ∎

From this we can easily deduce that chains of M adjacent nodal intervals all have approximately the same size, at least if n is large, and if we stay away from the endpoints. More precisely, we have the following result.

Corollary 2.5. *For any* $\eta > 0$ *and* $M \in \mathbb{N}$, *there is a* $K \in \mathbb{N}$ *so that if* $K \le k < k+j < n/2$, *then* $1 \le |I_{k+j}^n|/|I_k^n| < 1+\eta$ *whenever* $0 \le j < M$.

If nodal intervals were all the same size, then the distance between I_k^n and I_{k+j}^n would be exactly $(j-1)|I_k^n| = (j-1)|I_{j+k}^n|$. Because the nodal intervals vary in size, this is not true, but we do have the following similar estimate.

Lemma 2.6. *Suppose* $1 \le k < k + j < n/2$. *Then*

$$\operatorname{dist}(I_k^n, I_{k+j}^n) \ge \frac{2(j-1)(2k+j)}{n^2} \ge \frac{4}{\pi^2}(j-1)\Big(1 + \frac{j}{2k}\Big) \cdot |I_k^n|.$$

Proof. Let $c = \pi\frac{2k+1}{2n}$ be the right endpoint of J_k^n and $d = \pi\frac{2k+2j-1}{2n}$ the left endpoint of J_{k+j}^n. Then $-\cos(c)$ and $-\cos(d)$ are the right and left endpoints of I_k^n and I_{k+j}^n respectively, so

$$\operatorname{dist}(I_k^n, I_{k+j}^n) = \cos c - \cos d.$$

We can estimate this using the trigonometric identity

$$\cos c - \cos d = -2\sin\Big(\frac{c+d}{2}\Big)\sin\Big(\frac{c-d}{2}\Big) = 2\sin\Big(\frac{c+d}{2}\Big)\sin\Big(\frac{d-c}{2}\Big).$$

Recall that on $[0, \pi/2]$ we have $\sin x \ge 2x/\pi$. If $0 < c < d \le \pi/2$, then $(c+d)/2 \le \pi/2$ as well, so for $j \ge 1$ we get

$$\begin{aligned}
\cos c - \cos d &\ge 2 \cdot \frac{2}{\pi}\frac{c+d}{2} \cdot \frac{2}{\pi}\frac{d-c}{2} = \frac{2}{\pi^2}(d^2 - c^2) \\
&= \frac{1}{2}\Big(\frac{2k+2j-1}{n}\Big)^2 - \Big(\frac{2k+1}{n}\Big)^2 = \frac{1}{2}\frac{(2k+2j-1)^2 - (2k+1)^2}{n^2} \\
&= \frac{1}{2}\frac{(2j-2)(4k+2j)}{n^2} = 2\frac{(j-1)(2k+j)}{n^2}.
\end{aligned}$$

By Lemma 2.3, $|I_k^n| \le \pi^2 k/n^2$, so

$$\operatorname{dist}(I_k^n, I_{k+j}^n) \ge 2\frac{(j-1)(2k+j)}{n^2}\frac{|I_k^n|}{\pi^2 k/n^2} \ge \frac{4(j-1)(2k+j)}{2\pi^2 k}|I_k^n|. \tag{2.2}$$ ∎

Corollary 2.7. *Suppose* $1 \le k, k + j \le n$ *and* $j \ne 0$. *Then*

$$\frac{|I_k^n|}{\operatorname{dist}(I_k^n, I_{k+j}^n)} \le \frac{16}{|j|-1}.$$

Proof. Note that if I_k^n is farther from the origin than I_{j+k}^n is (or is equidistant), then all $|j|-1$ nodal intervals between them have length at least $|I_k^n|$, and therefore

$$\operatorname{dist}(I_k^n, I_{j+k}^n) \ge (|j|-1)|I_k^n|,$$

which is stronger than the inequality in the lemma.

Otherwise, I_k^n is strictly closer to the origin than I_{j+k}^n. First suppose they are on the same side of the origin. By symmetry, we may assume they are both to the left of the origin, so I_{k+j}^n is to the left of I_k^n, i.e., $j < 0$. In this case, (2.2) says

$$\frac{|I_k^n|}{\operatorname{dist}(I_k^n, I_{k+j}^n)} \le \frac{2\pi^2 k|I_k^n|}{2(|j|-1)(2k+|j|)|I_{j+k}^n|}.$$

Lemma 2.3 then gives

$$\frac{|I_k^n|}{\operatorname{dist}(I_k^n, I_{k+j}^n)} \le \frac{2\pi^2 k\,(1+\pi|j|/2k)}{2\,(|j|-1)\,(2k+|j|)} = \frac{\pi^2\,(k+(\pi/2)|j|)}{(|j|-1)\,(2k+|j|)}$$
$$\le \frac{\pi^2\,((\pi/2)k+(\pi/2)|j|)}{(|j|-1)(k+|j|)} = \frac{\pi^3}{2\,(|j|-1)} < \frac{16}{|j|-1},$$

as claimed in the lemma.

Finally, we must consider the case when I_k^n is strictly closer to the origin than I_{j+k}^n, but it is on the opposite side of the origin. Then I_{n-k}^n is between I_k^n and I_{k+j}^n, has the same size as I_k^n, and every interval between I_k^n and I_{n-k}^n is at least this long. There are $n-2k$ such intervals, including I_{n-k}^n but not I_k^n (the picture for n even and n odd is slightly different, but gives the same number in both cases). By symmetry, we may assume $k \le n/2$ and $j \ge 2n-k$. Thus

$$\operatorname{dist}(I_k^n, I_{k+j}^n) \ge (n-2k)|I_k^n| + \operatorname{dist}(I_{n-k}^n, I_j^n).$$

Now the previous case applies to the distance between I_{n-k}^n and I_{k+j}^n, and since $|I_k^n| = |I_{n-k}^n|$, we get

$$\operatorname{dist}(I_k^n, I_{k+j}^n) \ge (n-2k)\,|I_k^n| + \frac{1}{16}\,(j-(n-2k)-1)\,|I_{n-k}^n|$$
$$\ge \frac{1}{16}\,[(n-2k)\,|I_k^n| + (j-(n-2k)-1)\,|I_k^n|] = \frac{1}{16}\,(j-1)\,|I_k^n|. \quad \blacksquare$$

Numerical experiments suggest the corollary holds with the estimate $2/(|j|-1)$, although the particular value is not important for the proof of Theorem 1.1; any estimate of the form $O(1/(|j|-1))$ would work.

3. The area of a Chebyshev node

In this section, we estimate the integral of a Chebyshev polynomial T_n over a nodal interval I_k^n, and show that the result is approximately $(2/\pi)|I_k^n|$.

Lemma 3.1. *Suppose $J=[a,d]\subset[0,\pi]$ has length $t=d-a$ and $J'=[b,c]\subset J$ is concentric with length $s=c-b$. Then*

$$\frac{|\cos(J')|}{|\cos(J)|} \ge \frac{s}{t} = \frac{|J'|}{|J|}.$$

Proof. The intervals being concentric means that $(a+d)/2=(b+c)/2$. The difference formula for cosine implies

$$\frac{|\cos(J')|}{|\cos(J)|} = \frac{\cos b-\cos c}{\cos a-\cos d} = \frac{\sin\left(\frac{c+b}{2}\right)\sin\left(\frac{b-c}{2}\right)}{\sin\left(\frac{a+d}{2}\right)\sin\left(\frac{a-d}{2}\right)} = \frac{\sin\left(\frac{c-b}{2}\right)}{\sin\left(\frac{d-a}{2}\right)} = \frac{\sin(s/2)}{\sin(t/2)}, \tag{3.1}$$

so the claim is equivalent to whether for $0 \le s \le t \le \pi$ we have

$$\frac{\sin(s/2)}{\sin(t/2)} \ge \frac{s}{t}.$$

However, this is true because $\sin(x)/x$ is a decreasing function on $[0, \pi]$, as can be checked by differentiation. ∎

Lemma 3.2. *We have*

$$\int_{I_k^n} |T_n| \ge \frac{2}{\pi} |I_k^n|.$$

Proof. We will use the standard formula (see, e.g., Proposition 6.24 of [10])

$$\int_I f(x)\,dx = \int_0^\infty |\{x \in I : f(x) > t\}|\,dt, \tag{3.2}$$

that is valid for continuous, non-negative functions.

Recall that for $k = 1, \dots, n-1$, we set $J_k^n = [\pi \frac{2k-1}{2n}, \pi \frac{2k+1}{2n}]$, and we defined the nodal intervals $I_k^n = \cos(J_k^n)$. By the definition of the Chebyshev polynomials, if $x = \cos(y)$, then $T_n(x) = \cos(ny)$. Thus the interval $\{x \in I_k^n : T_n(x) > t\}$ is the image under cosine of the interval $J' = \{y \in J_k^n : \cos(ny) > t\}$. Please also note that it is without loss of generality to assume that T_n is positive on I_k^n.

We now apply Lemma 3.1 with $J = J_k^n$, $\cos(J) = I_k^n$, $J' = \{y \in J_k^n : |\cos(ny)| > t\}$ and $\cos(J') = \{x \in I_k^n : |T_n(x)| > t\}$. The function $|\cos(ny)|$ takes its maximum on J_k^n at the midpoint of J_k^n, and is symmetric with respect to this midpoint. Thus J' is concentric with J. Since $|J_k^n| = \pi/n$, by Lemma 3.1 we have

$$\frac{|\{x \in I_k^n : T_n(x) > t\}|}{|I_k^n|} \ge \frac{|\{y \in J_k^n : \cos(ny) > t\}|}{|J_k^n|} = \frac{n}{\pi} |\{y \in J : \cos(ny) > t\}|.$$

Using (3.2) twice gives

$$\begin{aligned}\frac{1}{|I_k^n|} \int_{I_k^n} T_n(x)\,dx &= \frac{1}{|I_k^n|} \int_0^1 |\{T_n(x) > t\}|\,dt \ge \frac{n}{\pi} \int_0^1 |\{\cos(ny) > t\}|\,dt \\ &= \frac{n}{\pi} \int_{J_k^n} \cos(ny)\,dy = \frac{n}{\pi}\frac{2}{\pi} \cdot |J_k^n| = \frac{n}{\pi}\frac{2}{\pi} \cdot \frac{\pi}{n} = \frac{2}{\pi}.\end{aligned}$$ ∎

A result of Erdős and Grünwald [8][1] gives a nearby upper bound.

Proposition 3.3. *If p is a polynomial with only real zeros and $a < b$ are roots of p with $p > 0$ on (a, b), then $\int_I p\,dx \le \frac{2}{3} |I| \max_I p$.*

Combining this with the previous lemma gives

$$.6366 \approx \frac{2}{\pi} \le \frac{\int_{I_n^k} |T_n(x)|\,dx}{|I_k^n|} \le \frac{2}{3} \approx .6666.$$

[1] This 1939 paper was one of the first listed in Mathematical Reviews: it has MR-number 7.

In our situation, we can improve this further. A direct calculation shows the integral actually converges to the lower bound as n increases (i.e., Chebyshev nodes "look like" nodes of sine).

Lemma 3.4. *With notation as above,*

$$\frac{2}{\pi} \leq \frac{\int_{I_k^n} |T_n|}{|I_k^n|} \leq \frac{2}{\pi} + \frac{\pi}{6n^2}.$$

Proof. Fix $0 < r < 1$, and as above set $J = J_k^n$ and $J' = \{y \in J_k^n : \cos(ny) > r\}$, $0 < r < 1$. Then $\cos(J) = I_k^n$ and $\cos(J') = \{x \in I_k^n : |T_n(x)| > r\}$. Set $s = |J'|$ and $t = |J|$ and note that $0 < s < t < \pi/n$. Using (3.1) we have

$$\begin{aligned}
\frac{|\{x \in I_k^n : |T_n(x)| > r\}|}{|I_k^n|} &= \frac{|\cos(J')|}{|\cos(J)|} = \frac{|J'|}{|J|} \cdot \frac{|J|}{|J'|} \cdot \frac{|\cos(J')|}{|\cos(J)|} \\
&= \frac{|\{y \in J_k^n : |\cos(ny)| > r\}|}{|J_k^n|} \cdot \frac{t}{s} \cdot \frac{\sin(s/2)}{\sin(t/2)} \\
&\leq \frac{|\{y \in J_k^n : |\cos(ny)| > r\}|}{|J_k^n|} \cdot \sup_{0<s<t<\pi/n} \frac{t \sin(s/2)}{s \sin(t/2)}.
\end{aligned}$$

We have already seen that

$$\frac{1}{|J_k^n|} \int_0^1 |\{y \in J_k^n : |\cos(ny)| > r\}|\, dr = \frac{1}{|J_k^n|} \int_{J_k^n} \cos(ny) dy = \frac{2}{\pi}.$$

Using this, (3.2), and $x - x^3/6 \leq \sin x \leq x$ on $[0, \pi/2]$, we can deduce that

$$\begin{aligned}
\frac{\int_{I_k^n} |T_n|}{|I_k^n|} &\leq \frac{2}{\pi} \cdot \sup_{0<s<t<\pi/2n} \frac{t \sin s}{s \sin t} \leq \frac{2}{\pi} \cdot \sup_{0<s<t<\pi/2n} \frac{ts}{s(t - t^3/6)} \\
&\leq \frac{2}{\pi} \cdot \sup_{0<s<t<\pi/2n} \frac{1}{1 - t^2/6}.
\end{aligned}$$

This is maximized at $t = \pi/2n$. Thus the last line above is less than

$$\frac{2}{\pi} \cdot \frac{1}{1 - (\pi/2n)^2/6} = \frac{2}{\pi} \cdot \frac{1}{1 - \pi^2/24n^2} \leq \frac{2}{\pi} \cdot \left(1 + \frac{\pi^2}{12n^2}\right) = \frac{2}{\pi} + \frac{\pi}{6n^2},$$

where the inequality holds because $1/(1 - x) \leq 1 + 2x$ for $0 \leq x \leq 1/2$, and since $\pi^2/(24n^2) < 1/2$ for $n \geq 1$. ■

Because T_n has opposite signs on adjacent nodal intervals, the integrals over adjacent intervals mostly cancel, especially when the intervals are close in length. Lemma 3.4 immediately implies the following estimate capturing this.

Corollary 3.5. *If I and J are adjacent nodal intervals with $|J| \leq |I| \leq (1 + \eta)|J|$, then*

$$\left|\frac{1}{|I| + |J|} \int_{I \cup J} T_n(x)\, dx\right| \leq \frac{2}{\pi} \cdot \frac{|I| - |J|}{|I| + |J|} + \frac{\pi}{6n^2} \leq \frac{\eta}{\pi} + \frac{\pi}{6n^2} < \frac{\eta}{3},$$

the last inequality holding for all $n \geq 6/\sqrt{\eta}$.

The argument proving Lemma 3.4 applies if $|T_n|$ is replaced by $h(|T_n|)$ for some increasing, continuous function h on $[0,1]$. For example, if $h(x) = x^2$, we deduce that

$$\frac{1}{2} \le \frac{1}{\pi}\int_0^\pi \sin^2(t)\,dt \le \int_{I_k^n} |T_n|^2(t)\,dt \le \frac{1}{2}\left(1+\frac{\pi^2}{12n^2}\right) = \frac{1}{2} + \frac{\pi^2}{24n^2}. \tag{3.3}$$

This estimate is utilized in [1].

4. Perturbing the roots

In this section, we introduce notation used in the rest of the paper, by defining the subintervals $\{G_k^n\} \subset [-1,1]$ formed by unions of adjacent nodal intervals, and the polynomials $T_n(x,y)$ created by moving the roots of $T_n(x)$ within these intervals.

For convenience, we will assume that n is of the form $n = 8m+1$ for some positive integer m. We do this because we are going to group the nodal intervals into groups of four to form intervals G_k^n, and we want the origin to be the common endpoint of two of these intervals. Thus the number of nodal intervals (which is $n-1$), must be a multiple of eight, leading to the condition $n = 8m+1$.

Fix such an n and let $T_n(x)$ be the nth Chebyshev polynomial. Let $N = (n-1)/4$. Define G_m^n as the union of the four nodal intervals

$$G_k^n = I_{4k-3}^n \cup I_{4k-2}^n \cup I_{4k-1}^n \cup I_{4k}^n = [r_{4k-3}^n, r_{4k+1}^n], \quad k = 1,\dots,N.$$

Each G_k^n has roots of T_n as endpoints, and it contains three other roots, r_{4k-2}^n, r_{4k-1}^n, and r_{4k}^n, in its interior. We will refer to these as the "interior roots" of G_k^n, and label them as "left", "center" and "right" respectively. Since n is odd, T_n is positive on the intervals I_k^n with k odd and is negative when k is even. Thus on each G_k^n the nodes are positive, negative, positive and negative moving left to right.

For each G_k^n, we will perturb the three interior roots and leave the endpoint roots of every G_k^n fixed. We describe the idea roughly here and more precisely in Section 6. Given a value $y \in [0,1/2]$, moving the roots by a factor of y will mean that the rightmost of the three interior roots is moved by distance $y \cdot |I_{4k-1}^n|$ to the right; if $y < 0$, then this root is moved distance $|y| \cdot |I_{4k-1}^n|$ to the left. We can describe both cases at once by saying the root is moved by the signed distance $y|I_{4k-1}^n|$. The middle and left interior roots are moved by signed distances $(-y-\delta)|I_{4k-1}^n|$ and $\delta|I_{4k-1}^n|$ respectively, where δ will be defined precisely in Section 6. Note that this causes the center of mass of the three roots to remained unchanged. In all the cases of interest, δ has the same sign as y, δ is comparable to $|y|$, and δ depends only on y and the ratio $a = |I_{4k-2}^n|/|I_{4k-1}^n|$, which will be close to 1. More concisely, $\delta = \delta(y,a) \simeq y$. The exact value of δ is chosen so that the perturbation of the three roots simultaneously has the effect of multiplying the polynomial by a rational function R so that $R(x) = 1 + O(|x|^{-3})$ as we move away from the perturbations. This decay will be verified in Section 6 and used in Sections 8 and 9 to show that the effect of distant perturbations is quite small compared to local ones.

In order to specify small perturbations of the Chebyshev roots in many different intervals at once, we introduce some notation. Let $Q_t^N = [-t,t]^n$ and let $\|(x_1,\dots,x_n)\| =$

$\max_{1 \le k \le n} |x_k|$ denote the supremum norm on $\mathbb{R}^N$. For a vector $y = (y_1, \ldots, y_N) \in Q_t^N$, let $\tilde{y}_m = (0, \ldots, y_m, \ldots, 0) \in Q_t^N$ be the vector that equals y_m in the mth coordinate and is zero in the other coordinates (i.e., the projection of y onto the mth coordinate axis). For $y \in Q_t^N$ and $x \in [-1, 1]$, define $T_n(x, y)$ to be the polynomial obtained from $T_n(x)$ by perturbing the zeros of T_n in G_k^n by a factor of y_k. When y is the all zeros vector, we make no perturbations, so $T_n(x, 0) = T_n(x)$. We can give a formula for the perturbed polynomial, although it is a bit awkward:

$$T_n(x, y) = 2^{n-1} \prod_{k=1}^{n} (x - z_k^n),$$

where

$$z_k^n = \begin{cases} r_k^n, \text{ if } k = 1 \bmod 4 \quad \text{(endpoints)}, \\ r_k^n + y_k |I_k^n|, \text{ if } k = 0 \bmod 4 \quad \text{(right interior)}, \\ r_k^n + \delta(y_k, |I_k^n|/|I_{k+1}^n|)|I_k^n|, \text{ if } k = 2 \bmod 4 \quad \text{(left interior)}, \\ r_k^n - (y_k + \delta(y_k, |I_k^n|/|I_{k+1}^n|))|I_k^n|, \text{ if } k = 3 \bmod 4 \quad \text{(center interior)}. \end{cases}$$

The perturbation is easier to describe in words than in this formula: each endpoint of every G_k^n is left fixed ($k = 1 \bmod 4$); the rightmost interior root ($k = 0 \bmod 4$) is moved by y_k times the length of the third component sub-interval $I = I_{4k-1}^n$; the leftmost root ($k = 2 \bmod 4$) is moved in the same direction by an amount $\delta|I|$, where δ depends on y_k and the length ratio of the center two intervals; the center root ($k = 3 \bmod 4$) moves in the opposite direction, and so that the center of mass of the three roots remains unchanged.

Define $A_k^n(y)$ as the average of $T_n(x, y)$ over G_k^n, i.e.,

$$A_k^n(y) = \frac{1}{|G_k^n|} \int_{G_k^n} T_n(x, y)\, dx, \tag{4.1}$$

for $k = 1, \ldots, N = (n-1)/4$. The proofs of the desired estimates for A_k^n will require that the four nodal intervals making up G_k^n to have nearly the same length, to within a fixed factor $1 + \eta$. By Corollary 2.5, we know this holds if we omit a finite number, K, of nodal intervals near each of -1 and 1. Thus we will study A_k^n only in the range $K < k \le N - K$ (depending on η, but not on n) of nodal intervals near each of -1 and 1. The number K depends on η, but not on n.

Recall $y = (y_1, \ldots, y_N)$ and $\tilde{y}_k = y_k$ in the kth coordinate and is zero elsewhere. We define maps f and g from Q_t^N into $\mathbb{R}^N$ whose coordinate functions for $K < k \le N - K$ are given by

$$f_k(y) = A_k^n(\tilde{y}_k) \tag{4.2}$$

and

$$g_k(y) = A_k^n(y). \tag{4.3}$$

Note that each f_k considers the integral of different polynomials on different intervals G_k^n (the polynomial using perturbations only in G_k^n), whereas g_k is defined using the same polynomial on every interval G_k^n.

In the remaining dimensions ($k \le K$ and $k > N - K$), we simply let f_k and g_k be the identity maps. These $2K$ dimensions play no role in the proof, and we might just define f and g as maps from Q_t^{N-2K} into $\mathbb{R}^{N-2K}$. However, this complicates the notation, and the indices of the coordinates of f and g would no longer match the indices of the intervals G_k^n, causing further confusion and inconvenience. Little is lost if the reader simply thinks of (4.2) and (4.3) as holding for all $1 \le k \le N$, but only verifies the arguments in the following sections when k is not too small or too large.

Each coordinate function f_k of f is real valued and only depends on y_k, the kth coordinate of y. Thus we can think of these maps as sending intervals to intervals. We will prove that each coordinate function f_k is monotone in y_k and that it maps $I_t = [-t, t]$ to a strictly larger interval; see Corollary 8.2. It follows that f is a homeomorphism of Q_t^N to some cube $Q' \supset Q_t^N$. In particular, $Q_{t/2}^n \subset f(Q_t^n)$. We want to show the same is true for g, i.e., $Q_{t/2}^N \subset g(Q_t^N)$. This will imply that given a Lipschitz function F, we can find a perturbed Chebyshev polynomial with an anti-derivative P_n that agrees with F at the endpoints of every $\{G_k^n\}$. Since both F and P_n are Lipschitz, and the lengths of G_k^n tend to zero uniformly with n, this will imply that P_n converges uniformly to F. See Theorem 10.2.

The following result is a precise formulation of the idea that the integral of the perturbed polynomial $T_n(x, y)$ over the interval G_k^n is dominated by the perturbation of the roots inside G_k^n, and that the perturbations exterior to G_k^n have a strictly smaller effect.

Theorem 4.1. *There exists $t > 0$ such that the following holds. If n and K are large enough, and the maps f and g are defined as above for $y \in Q_t^N$, then $Q_t^N \subset f_k(Q_t^N)$ and $\|f_k - g_k\|_{Q_t^N} \le t/2$ for $1 \le k \le N$.*

This will be proven in Sections 6 to 9. If N were equal to 1, then Theorem 4.1 and the intermediate value theorem would immediately imply that $g(Q_t^1)$ covers $Q_{t/2}^1$. In Section 10, we will use Brouwer's theorem to draw the same conclusion in the higher dimensional case.

5. The distortion of 2-point perturbations

Before discussing 3-point perturbations, we briefly discuss moving just two roots. It seems worthwhile to do this, since the 3-point perturbation can be thought of as the composition of two 2-point perturbations, and the discussion of the 2-point perturbation explains the main idea in a simpler setting.

If a polynomial has zeros at ± 1 and we move these by ε to the left and right respectively, this is the same as multiplying the polynomial by (see Figure 6)

$$R(x) = \frac{(x - 1 - \varepsilon)(x + 1 + \varepsilon)}{(x - 1)(x + 1)} = \frac{x^2 - (1 + \varepsilon)^2}{x^2 - 1} = 1 - \frac{2\varepsilon + \varepsilon^2}{x^2 - 1}. \tag{5.1}$$

Then $R(x) \ge 1 + 2\varepsilon + \varepsilon^2$ on $[-1, 1]$, and $0 < R(x) < 1$ on $\{x : |x| > 1 + \varepsilon\}$.

If we have a polynomial p that has roots at ± 1, and we move these roots to $\pm(1 + \varepsilon)$, then the calculation above says that the new polynomial $\widetilde{p}$ is larger than p (in absolute value) in $[-1, 1]$ and is smaller (in absolute value) outside $[-1 - \varepsilon, 1 + \varepsilon]$. We call the

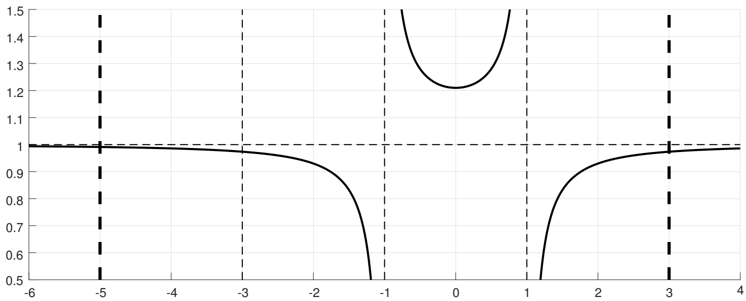

Figure 6. A plot of $r(t) = (a+1+\varepsilon)(x-1-\varepsilon)/(x^2-1)$. If P has roots at ± 1, and we move them by $\pm\varepsilon$, the new polynomial is $\widetilde{p} = R \cdot p$.

ratio $R(x)$ the 2-point distortion function, since it describes how a polynomial is altered by moving two of its roots.

We want to generalize this to a polynomial p with roots at $\{a, b\}$, that we move slightly to $\{a-(b-a)\varepsilon/2, b+(b-a)\varepsilon/2\}$ (each root gets moved by ε, relative to the length of the interval). We say the roots have been "perturbed by a factor of ε". Moving the roots apart in this way multiplies p by at least $1+2\varepsilon+\varepsilon^2$ in the interval I, and decreases it outside I in absolute value by a factor

$$1 - \frac{2\varepsilon+\varepsilon^2}{4(\operatorname{dist}(x,c)/|I|)^2 - 1}, \tag{5.2}$$

where $c = (b+a)/2$ is the center of the interval. An example of perturbing two roots of a Chebyshev polynomial was illustrated in Figure 1.

6. The distortion of 3-point perturbations

In this section, we describe the distortion caused by moving three adjacent roots of a Chebyshev polynomial. The two nodal intervals with these endpoints need not have equal lengths, but we may assume they have length ratio close to 1, and we will model the situation using three points $\{-a, 0, 1\}$, where $a = 1+\alpha$, with α small, say $|\alpha| < 1/10$. After rescaling, this will cover all the cases that are needed later.

Suppose we have a polynomial p with roots at $\{-a, 0, 1\}$, among possibly many other roots. We create a new polynomial $\widetilde{p}$ by moving these three roots to $-a+\delta$, $-\varepsilon-\delta$ and $1+\varepsilon$, respectively. Note that this keeps the center of mass of the roots unchanged at $(1-a)/3$. See Figure 7.

We can think of this 3-point perturbation as the composition of two 2-point perturbations: first moving the pair $\{0, 1\}$ to $\{-\varepsilon, 1+\varepsilon\}$, and then moving the pair $\{-a, -\varepsilon\}$ to $\{-a+\delta, -\varepsilon-\delta\}$. Since each move changes p by a factor of $1+O(x^{-2})$ far from the origin, the combined motion also has at least this decay rate. However, by carefully choosing δ (depending on a and ε), we can arrange for cancelation that improves the decay rate to $O(|x|^{-3})$. The remainder of this section explains how to do this.

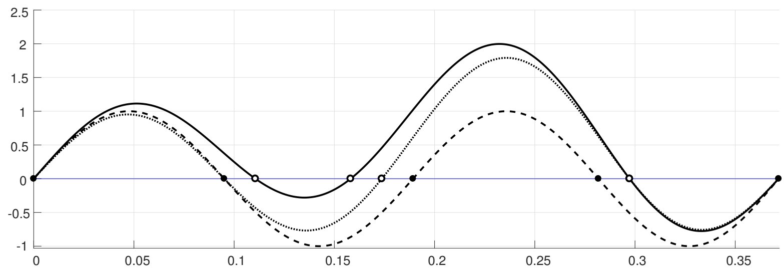

Figure 7. The original Chebyshev polynomial is solid, the 2-point perturbation is dotted and the 3-point perturbation is dashed. Each perturbation is strictly larger than the unperturbed polynomial over the nodal intervals adjacent to the perturbed roots, but this need not hold further away: the 2-point perturbation is slightly smaller than T_n in the leftmost nodal interval.

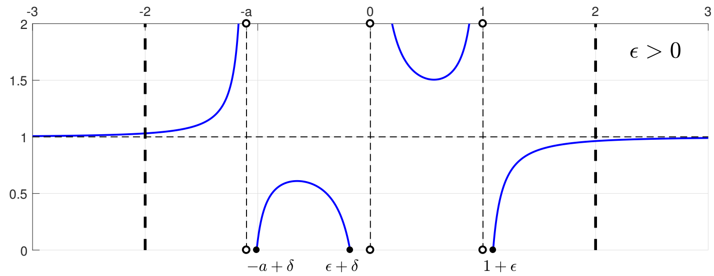

Figure 8. The rational function R corresponding to the 3-point perturbation $\{-a, 0, 1\} \to \{-a + \delta, -\delta - \varepsilon, 1 + \varepsilon\}$. Here we have taken $a = 1.1$ and $\delta = \varepsilon = .09 > 0$. Over $[0, 1]$, R is bounded strictly above 1 by an amount comparable to the size of the perturbation, and on the other three nodal intervals, the perturbed polynomial is larger than the original.

Note that $\widetilde{p} = p \cdot R$ where

$$R(x) = \frac{P(x)}{Q(x)} := \frac{(x + a - \delta)(x + \delta + \varepsilon)(x - 1 - \varepsilon)}{(x + a)\,x\,(x - 1)}.$$

We call R the 3-point distortion function associated with the perturbation. See Figure 8. We can write

$$P(x) = x^3 + Ax^2 + Bx + C,$$

where

$$\begin{aligned}
A &= (a - \delta) + (\delta + \varepsilon) + (-1 - \varepsilon) = a - 1,\\
B &= (a - \delta)(\delta + \varepsilon) + (a - \delta)(-1 - \varepsilon) + (\delta + \varepsilon)(-1 - \varepsilon)\\
&= (a\delta + a\varepsilon - \delta^2 - \delta\varepsilon) + (-a - a\varepsilon + \delta + \delta\varepsilon) + (-\delta - \delta\varepsilon - \varepsilon - \varepsilon^2)\\
&= -a - \varepsilon + a\delta - \delta\varepsilon - \delta^2 - \varepsilon^2,\\
C &= (a - \delta)(\delta + \varepsilon)(-1 - \varepsilon) = (a\delta + a\varepsilon - \delta^2 - \delta\varepsilon)(-1 - \varepsilon)\\
&= -a\delta - a\varepsilon + \delta^2 + (1 - a)\delta\varepsilon - a\varepsilon^2 + \delta^2\varepsilon + \delta\varepsilon^2.
\end{aligned}$$

Then

$$\begin{aligned} R(x) = \frac{P(x)}{Q(x)} &= 1 + \frac{P(x)-Q(x)}{Q(x)} \\ &= 1 + \frac{(x^3+Ax^2+Bx+C)-(x^3+(a-1)x^2-ax)}{(x+a)x(x-1)} \\ &= 1 + \frac{(A-a+1)x^2+(B+a)x+C)}{(x+a)x(x-1)} = 1 + \frac{bx+C}{(x+a)x(x-1)}, \end{aligned}$$

where

$$b = B + a = a\delta - \varepsilon - \delta\varepsilon - \delta^2 - \varepsilon^2.$$

We want to choose δ so that $b = 0$. If $\delta = \lambda\varepsilon$, then the equation $b = 0$ becomes

$$\begin{aligned} 0 = a\lambda\varepsilon - \varepsilon - \lambda\varepsilon^2 - \lambda^2\varepsilon^2 - \varepsilon^2 \quad &\Longrightarrow \quad 0 = a\lambda - 1 - \varepsilon(\lambda^2+\lambda+1) \\ &\Longrightarrow \quad \varepsilon = \frac{a\lambda - 1}{\lambda^2+\lambda+1} = r_a(\lambda). \end{aligned}$$

The rational function r_a on the right has a zero at $1/a$ where it has slope $\approx 1/3$. More precisely, if $a = 1+\alpha$, then using long division of polynomials we get

$$r_a'\left(\frac{1}{a}\right) = \frac{a^3}{1+a+a^2} = \frac{1}{3}\cdot\frac{1+3\alpha+3\alpha^2+\alpha^3}{1+\alpha+\alpha^2/3} = \frac{1}{3}\Big[1+2\alpha+\frac{2}{3}\alpha^2+O(\alpha^3)\Big].$$

So if ε is small and $a \approx 1$ (both will hold in our application), then the equation $r_a(\lambda) = \varepsilon$ will have a solution of the form

$$\lambda = (1/a) + 3\varepsilon/a + O(\varepsilon^2 + \varepsilon\alpha^2).$$

Hence $\delta = \varepsilon/a + O(\varepsilon^2)$. See Figure 9. The calculations up to this point prove the following result, giving the desired cubic decay of the distortion function R.

Lemma 6.1. *If $|a-1| < 1/5$ and $0 \le \varepsilon \le 1/10$, then we can make a 3-point perturbation of the form $\{-a, 0, 1\} \to \{-a+\delta, -\delta-\varepsilon, 1+\varepsilon\}$ so that $\delta = \varepsilon/a + O(\varepsilon^2)$, and the distortion equals*

$$R(x) = 1 + \frac{C}{Q(x)} = 1 + \frac{C}{(x+a)x(x-1)}, \tag{6.1}$$

where

$$C = -a\delta - a\varepsilon + \delta^2 + (1-a)\delta\varepsilon - a\varepsilon^2 + \delta^2\varepsilon + \delta\varepsilon^2 = -(1+a)\varepsilon + O(\varepsilon^2). \tag{6.2}$$

The following lemma implies the perturbed polynomial moves monotonically as a function of the perturbation parameter ε. See Figures 10 and 12. This will later imply that the function f defined in (4.2) is a homeomorphism, which will be needed in our application of Brouwer's theorem (see Lemma 10.1).

Lemma 6.2. *Suppose that $\widetilde{p}_1$ and $\widetilde{p}_2$ are perturbations of $p(x) = (x+a)x(x-1)$ as described above, by factors of $\varepsilon_1 < \varepsilon_2$ respectively. Then $\widetilde{p}_1 > \widetilde{p}_2$ for all x.*

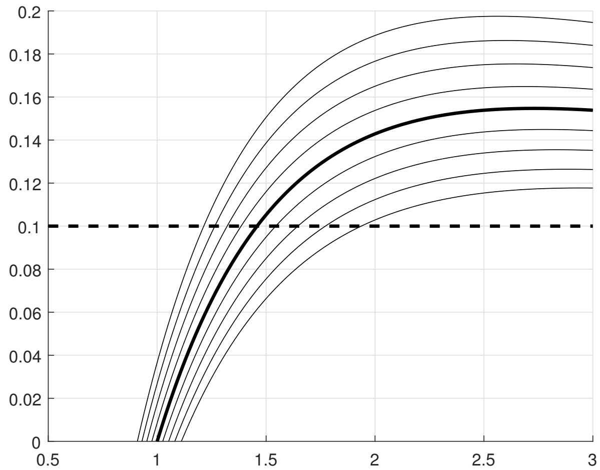

Figure 9. A plot of $r(t) = (at - 1)/(1 + t + t^2)$ for $a = \{.8, .85, \dots 1.2\}$. The curve for $a = 1$ is thickened. The plots show that $r(t) = \varepsilon$ has a solution for all $\varepsilon \in [0, .1]$ and all $a \in [.8, 1.2]$.

Proof. Note that

$$\widetilde{p}_1 - \widetilde{p}_2 = R_1 p - R_2 p = \left(1 + \frac{C_1}{Q(x)}\right) Q(X) - \left(1 + \frac{C_2}{Q(x)}\right) Q(X) = C_1 - C_2.$$

Thus $\widetilde{p}_1 > \widetilde{p}_2$ if and only if $C_1 > C_2$. Thus we either have $\widetilde{p}_1 > \widetilde{p}_2$ everywhere, or $\widetilde{p}_1 \le \widetilde{p}_2$ everywhere. Since $\varepsilon_1 < \varepsilon_2$, the rightmost perturbed root of $\widetilde{p}_1$ is to the left of the corresponding root y of $\widetilde{p}_2$, and hence $\widetilde{p}_1(y) > \widetilde{p}_2(y) = 0$. Thus $\widetilde{p}_1 > \widetilde{p}_2$ everywhere. ■

Note that in (6.2), if ε is small and a is close to 1 (which is the case in our applications), then we have $C \approx -2\varepsilon$. In particular, for ε sufficiently small, $R(x) - 1$ has the opposite sign as $Q(x) = (x + a)x(x - 1)$. Therefore we get the following inequalities. See Figure 8.

Corollary 6.3. *For $|\varepsilon|$ small enough,*

$$\begin{aligned}
R(x) &\ge 1, \quad \textit{if } \varepsilon > 0 \textit{ and } x \in (-\infty, -a] \cup [0, 1],\\
R(x) &\ge 1, \quad \textit{if } \varepsilon < 0 \textit{ and } x \in (-a, 0] \cup [1, \infty),\\
R(x) &\le 1, \quad \textit{if } \varepsilon < 0 \textit{ and } x \in (-\infty, -a] \cup [0, 1],\\
R(x) &\le 1, \quad \textit{if } \varepsilon > 0 \textit{ and } x \in (-a, 0] \cup [1, \infty).
\end{aligned}$$

We shall give separate estimates for $R(x)$ on different sets of intervals based on their distance to the origin: we call these cases "near" ($[-a, 1]$), "intermediate" ($[-2, -a]$ and $[1, 2]$) and "far" ($\{x : |x| \ge 2\}$). Since $1 + a \approx 2$ if $a \approx 1$ and $|(x + a)x(x - 1)| \approx 6$ at $x = \pm 2$, we can immediately deduce the following for the intermediate intervals.

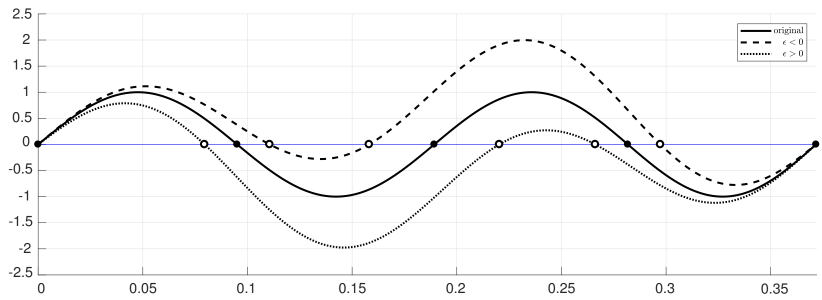

Figure 10. A 3-point perturbation for both positive and negative values of ε. The solid graph is the unperturbed Chebyshev polynomial over one group of four nodal intervals. The dashed line is the $\varepsilon > 0$ perturbation and the dotted is the $\varepsilon < 0$ perturbation. This illustrates the monotonic movement proven in Lemma 6.2. A wider range of perturbations is shown in Figure 12.

Corollary 6.4. *For $|\varepsilon|$ small enough and $a = 1 + \alpha$ close to* 1*, we have*

$$R(x) \geq 1 + \left[\frac{1}{3} + O(\alpha) + O(\varepsilon)\right]\varepsilon > 1, \quad \text{if } \varepsilon > 0 \text{ and } x \in [-2, -a],$$
$$R(x) \leq 1 - \left[\frac{1}{3} + O(\alpha) + O(\varepsilon)\right]\varepsilon < 1, \quad \text{if } \varepsilon > 0 \text{ and } x \in [1, 2],$$
$$R(x) \leq 1 + \left[\frac{1}{3} + O(\alpha) + O(\varepsilon)\right]\varepsilon < 1, \quad \text{if } \varepsilon < 0 \text{ and } x \in [-2, -a],$$
$$R(x) \geq 1 - \left[\frac{1}{3} + O(\alpha) + O(\varepsilon)\right]\varepsilon > 1, \quad \text{if } \varepsilon < 0 \text{ and } x \in [1, 2].$$

Next we estimate the distortion $R(x)$ on the "far" intervals.

Corollary 6.5. *Suppose $a = 1 + \alpha$ with $|\alpha| < 1/5$. Then for $|x| \geq 2$,*

$$R(x) = 1 - \frac{[2 + O(\alpha) + O(\varepsilon)]\,\varepsilon}{(x+1)\,x\,(x-1)}.$$

Proof. By Lemma 6.1,

$$R(x) = 1 + \frac{-(1+a)\varepsilon + O(\varepsilon^2)}{(x+a)\,x\,(x-1)} = 1 - \frac{(1+a)\varepsilon + O(\varepsilon^2)}{(x+1)\,x\,(x-1)} \cdot \left(\frac{x+a+1-a}{x+a}\right)$$
$$= 1 - \frac{(1+a)\varepsilon + O(\varepsilon^2)}{(x+1)\,x\,(x-1)} \cdot \left(1 + \frac{1-a}{x+a}\right).$$

The maximum absolute value of $1 + (1-a)/(x+a)$ on $\{x : |x| \geq 2\}$ is attained at either $x = 2$ (if $a > 1$) or $x = -2$ (if $a < 1$), where the values are, respectively,

$$1 + \frac{1-a}{2+a} = \frac{3}{2+a} = 1 - \frac{\alpha}{3} + O(\alpha^2),$$
$$1 + \frac{1-a}{-2+a} = \frac{-1}{2-a} = -1 - \alpha + O(\alpha^2).$$

In either case, we deduce that the maximum absolute value is $1 + O(\alpha)$. Thus

$$R(x) = 1 - \frac{(2+\alpha)\varepsilon[1 + O(\alpha) + O(\varepsilon^2)]}{(x+1)x(x-1)} = 1 - \frac{(2 + O(\alpha))\varepsilon + O(\varepsilon^2)}{(x+1)x(x-1)}. \qquad \blacksquare$$

Next we consider the distortion near the perturbed points. There are separate estimates for each sub-interval $[0, 1]$ and $[-a, 0]$.

Lemma 6.6. *If* $\varepsilon > 0$, $a = 1 + \alpha$ *and* $x \in [0, 1]$, *then*

$$R(x) \geq 1 + \Big(4 - \frac{1}{a^2}\Big)\varepsilon + O(\varepsilon^2) \geq 1 + [3 + O(\alpha) + O(\varepsilon)]\varepsilon > 1.$$

Proof. For $x \in [0, 1]$, $R(x) \geq 1$, and we can write

$$\begin{aligned} R(x) &= \Big(1 - \frac{\delta}{x+a}\Big)\Big(1 + \frac{\varepsilon+\delta}{x}\Big)\Big(1 - \frac{\varepsilon}{x-1}\Big) \\ &= \Big(1 - \frac{\delta}{x+a}\Big)\Big(1 + \frac{\varepsilon+\delta}{x}\Big)\Big(1 + \frac{\varepsilon}{1-x}\Big), \end{aligned}$$

and all three terms in the last line are positive for $0 \leq x \leq 1$. Thus we make the second term smaller by subtracting the positive term δ/x, and so

$$R(x) \geq \Big(1 - \frac{\delta}{a}\Big)\Big(1 + \frac{\varepsilon}{x}\Big)\Big(1 + \frac{\varepsilon}{1-x}\Big).$$

By symmetry, the function on the right takes a minimum at the midpoint of the two poles, i.e., at $x = 1/2$ (one can also verify this by differentiating). Thus on $[0, 1]$, R is bounded below by

$$\begin{aligned} \Big(1 - \frac{\delta}{a}\Big)\Big(1 + \frac{\varepsilon}{1/2}\Big)\Big(1 + \frac{\varepsilon}{1-(1/2)}\Big) &= \Big(1 - \frac{\delta}{a}\Big)(1+2\varepsilon)(1+2\varepsilon) \\ &= (1 - \varepsilon a^{-2} + O(\varepsilon^2))(1 + 4\varepsilon + 4\varepsilon^2) = 1 + \Big(4 - \frac{1}{a^2}\Big)\varepsilon + O(\varepsilon^2). \qquad \blacksquare \end{aligned}$$

Lemma 6.7. *If* $\varepsilon > 0$, $a = 1 + \alpha$ *and* $x \in [-a, 0]$, *then*

$$R(x) \leq 1 - \Big(\frac{4}{a^2} - 1\Big)\varepsilon + O(\varepsilon^2) \leq 1 - (3 + O(\alpha) + O(\varepsilon))\varepsilon < 1.$$

Proof. For $x \in [-a, 0]$, $R(x) < 1$ and we have

$$R(x) = \Big(1 - \frac{\delta}{x+a}\Big)\Big(1 + \frac{\varepsilon+\delta}{x}\Big)\Big(1 - \frac{\varepsilon}{x-1}\Big) = \Big(1 - \frac{\delta}{x+a}\Big)\Big(1 + \frac{\varepsilon+\delta}{x}\Big)\Big(1 + \frac{\varepsilon}{1-x}\Big).$$

Since $\varepsilon > 0$, the last term in the product is ≥ 0 for $x \in [-a, 0]$, but the other two terms may be positive or negative.

For example, if the second term is negative, then $x \in (-\varepsilon - \delta, 0)$. In this case, the first term is approximately equal to $1 + \delta/a$, and since δ is positive if ε is positive, this term of the product must be negative. Therefore the whole product is negative in this case, so $R(x) < 0 < 1$ and the estimate in the lemma is certainly true.

If the first term in the product is negative, then $x \in (-a, -a+\delta)$, and the second term is approximately $1-(\varepsilon+\delta)/a$. This is positive since ε and δ are both small and $a \approx 1$. Again, the whole product must therefore be positive in this case, and the lemma holds. Thus we can restrict attention to the case when all three terms of the product are positive, i.e., $-1+\delta < x < -\varepsilon-\delta$.

In this case, the middle term of the product is made larger by removing the negative term δ/x, and the third term increases by setting $x=0$. Thus on $[-a,0]$ we have

$$R(x) \le \Big(1-\frac{\delta}{x+a}\Big)\Big(1+\frac{\delta}{x}\Big)(1+\varepsilon).$$

By symmetry, this function of x takes its maximum at the midpoint of the two poles, i.e., $x=-a/2$ (we can also check by direct calculation that the derivative of R is increasing on this interval and it is only zero at $x=-a/2$, so this point is the global minimum of R over this interval). Thus for $x \in (-a+\delta, -\varepsilon-\delta)$ we deduce that

$$\begin{aligned} R(x) &\le \Big(1-\frac{\delta}{a/2}\Big)\Big(1+\frac{\delta}{-a/2}\Big)(1+\varepsilon) \le \Big(1-\frac{2\varepsilon+O(\varepsilon^2)}{a^2}\Big)\Big(1-\frac{2\varepsilon+O(\varepsilon^2)}{a^2}\Big)(1-\varepsilon) \\ &\le \Big(1-\frac{2\varepsilon+O(\varepsilon^2)}{a^2}\Big)^2(1-\varepsilon) \le 1-\Big(\frac{4}{a^2}-1\Big)\varepsilon+O(\varepsilon^2). \qquad \blacksquare \end{aligned}$$

For $\varepsilon < 0$, the calculations are almost identical, just the logic of which terms are positive or negative changes. We state the corresponding estimates, but leave the verification to the reader.

Lemma 6.8. *If $\varepsilon < 0$, $a = 1+\alpha$ and $x \in [0,1]$, then*

$$R(x) \le 1+\Big(4-\frac{1}{a^2}\Big)\varepsilon+O(\varepsilon^2) = 1-(3+O(\alpha)+O(\varepsilon))\,|\varepsilon| < 1.$$

Lemma 6.9. *If $\varepsilon < 0$, $a = 1+\alpha$ and $x \in [-a,0]$, then*

$$R(x) \ge 1-\Big(\frac{4}{a^2}-1\Big)\varepsilon+O(\varepsilon^2) = 1+(3+O(\alpha)+O(\varepsilon))\,|\varepsilon| > 1.$$

These lemmas show that we can take $|R(x)-1| \ge \lambda|\varepsilon|$ on $[-a,1]$ for any $\lambda < 3$, by taking $|\varepsilon|$ sufficiently small, and a sufficiently close to 1.

We will not use the estimates derived above in precisely the form they were given. Instead, we will use rescaled versions, which we now state explicitly. Recall that we chose n of the form $n = 8m+1$, so that there were $n-1 = 8m = 2N$ nodal intervals $\{I_j^n\}_1^{n-1}$, and that we defined intervals $\{G_k^n\}_1^N$ by taking groups of four adjacent nodal intervals. More precisely, $G_k^n = I_{4k-3}^n \cup I_{4k-2}^n \cup I_{4k-1}^n \cup I_{4k}^n$. Moreover, since $I_k^n = [r_k^n, r_{k+1}^n]$, where $\{r_j^n\}_1^n$ are the roots of T_n, we have $G_k^n = [r_{4k-3}^n, r_{k+1}^n]$, and the three interior roots of G_k^n are r_{4k-2}^n, r_{4k-1}^n and r_{4k}^n. See Figure 11.

Define a linear map $\phi_k^n\colon \mathbb{R} \to \mathbb{R}$ by the conditions $\phi_k^n(r_{4k-1}^n) = 0$ and $\phi_k^n(r_{4k}^n) = 1$. In other words, we map the center and right interior roots of G_k^n to 0 and 1 respectively. By Corollary 2.5, if G_k^n is not too close to either -1 or 1, then the four nodal intervals it contains are all approximately the same size, and so we get

$$s = \phi_k^n(r_{4k-3}^n) \approx -2, \quad -a = \phi_k^n(r_{4k-2}^n) \approx -1, \quad t = \phi_k^n(r_{4k+1}^n) \approx 2.$$

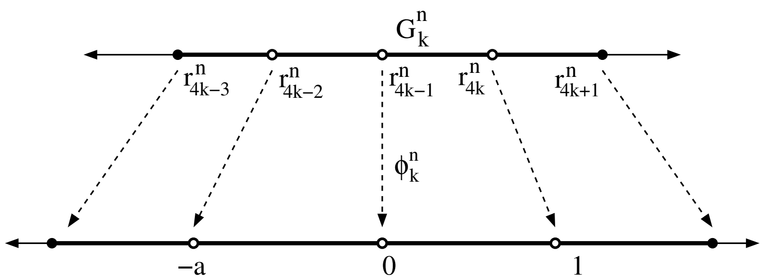

Figure 11. The definition of G^n_k and the linear map ϕ^n_k. The white dots are the three roots interior to G^n_k; these map to $-a, 0, 1$ under ϕ^n_k.

Using the map ϕ^n_k, a polynomial p with three roots inside G^n_k corresponds to polynomial q with three roots in $[s,t]$ by $p(x) = q(\phi^n_k(x))/|(\phi^n_k)'|^3$. If we take the ratio of two such polynomials p_1 and p_2, then the derivative factor cancels and we see that

$$\frac{p_1(x)}{p_2(x)} = \frac{q_1(\phi^n_k(x))}{q_2(\phi^n_k(x))}.$$

Thus the distortion function R^n_k for perturbations on G^n_k is just a linear rescaling of the 3-point distortion function R defined earlier in this section, i.e., $R^n_k(x) = R(\phi^n_k(x))$. With this, we can restate the results above for perturbations of roots in G^n_k. For example, the following are the rescaled versions of Corollaries 6.3 and 6.5.

Corollary 6.10. *If we perturb the interior roots of G^n_k by a factor of ε, and $|\varepsilon|$ is small enough, then*

$$\begin{aligned}
R(x) &\geq 1, \quad \textit{if } \varepsilon > 0 \textit{ and } x \in (-\infty, r^n_{4k-2}] \cup [r^n_{4k-1}, r^n_{4k}],\\
R(x) &\geq 1, \quad \textit{if } \varepsilon < 0 \textit{ and } x \in (r^n_{4k-2}, r^n_{4k-1}] \cup [r^n_{4k}, \infty),\\
R(x) &\leq 1, \quad \textit{if } \varepsilon < 0 \textit{ and } x \in (-\infty, r^n_{4k-2}] \cup [r^n_{4k-1}, r^n_{4k}],\\
R(x) &\leq 1, \quad \textit{if } \varepsilon > 0 \textit{ and } x \in (r^n_{4k-2}, r^n_{4k-1}] \cup [r^n_{4k}, \infty).
\end{aligned}$$

Corollary 6.11. *Suppose $a = 1 + \alpha$ with $|\alpha| < 1/5$. For $x \in [-1,1] \setminus G^n_k$, let*

$$d = \frac{|x - r^n_{4k-2}|}{|r^n_{4k-1} - r^n_{4k}|} = |x - r^n_{4k-2}| \cdot |(\phi^n_k)'|$$

be the distance between x and the center root of G^n_k, normalized by $|I^n_{4k-1}|$. Then

$$R(x) = 1 - \frac{[2 + O(\alpha) + O(\varepsilon)]\,\varepsilon}{(d+1)\,d\,(d-1)}.$$

Similarly, Lemmas 6.6 to 6.9 can be restated for perturbations of the interior roots of G^n_k by leaving the estimate for R exactly the same as before, and simply replacing the intervals for x by new intervals obtained by replacing the positions $\{-2, -a, 0, 1, 2\}$ by the points $\{r^n_{4k-3}, r^n_{4k-2}, r^n_{4k-1}, r^n_{4k}, r^n_{4k+1}\}$. The leftmost and rightmost points do not correspond exactly to -2 and 2 under ϕ^n_k, but only Corollary 6.4 makes use of these points, and in this case the corresponding image points are so close that the estimate still holds in the rescaled case.

7. Bounding the extreme values

In this section, we show that small perturbations of the roots do not increase the extreme values very much. This is needed in order to show that our polynomial approximants can be taken to be Lipschitz if the function f being approximated is Lipschitz (as claimed in Theorem 1.1). In a later section, we will also use these estimates to bound the size of the set where this perturbed Chebyshev polynomial is close to zero, in order to prove our approximants have derivatives that diverge pointwise almost everywhere. See Lemma 12.1 and Corollary 12.2.

We first prove a special case ($n = 3$) of the min-max property of the Chebyshev polynomials, that was mentioned in Section 2.

Lemma 7.1. *Suppose $r_1, r_2, r_3 \in I = [-1, 1]$ and let $p(x) = (x - r_1)(x - r_2)(x - r_3)$. Then $\max_I |p| \geq 1/4$ and the maximum is minimized by taking $r_1 = 0$ and $r_2, r_3 = \pm\sqrt{3}/2$ (in other words, p is a multiple of the Chebyshev polynomial T_3).*

Proof. A direct calculation shows that the given roots satisfy $\max_I |p| = 1/4$, so we only need to show that this is the best possible. If q minimizes the supremum of $|p|$ over I among cubic monic polynomials with roots in $[-1, 1]$ (a minimum exists by compactness), then let $\widetilde{q}(x) = \frac{1}{2}(q(x) - q(-x))$. This is also cubic, monic, and satisfies

$$\sup_I \left| \frac{q(x) - q(-x)}{2} \right| \leq \sup_I |q(x)|.$$

This polynomial is clearly odd, so it has a root at 0. If the other two roots were complex, they would have to be both complex conjugates of each other and also negatives of each other, and hence both would be zero. In this case $\widetilde{q}(x) = x^3$ and $\widetilde{q}(1) = 1 > 1/4$, so this is not the minimum.

Thus the minimizing polynomial q is odd with only real roots: $q(x) = x(x^2 - r^2)$ for some $0 < r \leq 1$. If $r < \sqrt{3}/2$, then $q(1) = (1 - r)(1 + r) = 1 - r^2 > 1 - 3/4 = 1/4$, so q is not minimizing. If $r > \sqrt{3}/2$, then

$$q\Big(\frac{1}{2}\Big) < \frac{1}{2}\Big(\frac{1}{2} - \frac{\sqrt{3}}{2}\Big)\Big(\frac{1}{2} + \frac{\sqrt{3}}{2}\Big) = \frac{1}{2}\Big(\frac{1}{4} - \frac{3}{4}\Big) = -\frac{1}{4}.$$

This is not minimizing either, so the optimal r equals $\sqrt{3}/2$, as claimed. ∎

Lemma 7.2. *For any $\lambda > 1$, there is a $\varepsilon > 0$ so that, if $\|y\| \leq \varepsilon$ and $T_n(x, y)$ is the corresponding perturbation of the Chebyshev polynomial T_n, then*

$$\frac{1}{\lambda} \sup_{G_m^n} |T_n(x)| \leq \sup_{G_m^n} |T_n(x, y)| \leq \lambda \sup_{G_m^n} |T_n(x)|.$$

Proof. By Corollary 6.11, the perturbations performed outside G_m^n only multiply T_n by a factor of $1 + O(\varepsilon)$ inside G_m^n, so it suffices to show that the perturbations inside G_m^n only change the supremum by a similar factor.

For brevity, let $I = G_m^n$. Suppose we have three points $r_1, r_2, r_3 \in I$ and consider $h(x) = (x - r_1)(x - r_2)(x - r_3)$. By rescaling $[-1, 1]$ to I in Lemma 7.1, the supremum of $|h|$ over I is minimized (over choices of the roots in I) by taking the degree 3 Chebyshev polynomial on I, and in this case the minimum is $(1/4)(|I|/2)^3 = |I|^3/32$. Thus $\sup_I h \geq |I|^3/32$.

Fix $\varepsilon > 0$ and let $|\varepsilon_j| \leq \varepsilon|I|$ for $j = 1, 2, 3$. Define the following perturbation of h:

$$\widetilde{h}(x) = (x - r_1 + \varepsilon_1)(x - r_2 + \varepsilon_2)(x - r_3 + \varepsilon_3).$$

Then if $\varepsilon < 1/2$,

$$\begin{aligned} |h(x) - \widetilde{h}(x)| &= |(x-r_1)(x-r_2)(x-r_3) - (x-r_1+\varepsilon_1)(x-r_2+\varepsilon_2)(x-r_3+\varepsilon_3)| \\ &\leq |x-r_1||x-r_2||\varepsilon_3| + |x-r_1||x-r_3||\varepsilon_2| + |x-r_2||x-r_3||\varepsilon_1| \\ &\quad + |x-r_1||\varepsilon_2||\varepsilon_3| + |x-r_2||\varepsilon_1||\varepsilon_3| + |x-r_3||\varepsilon_1||\varepsilon_2| + |\varepsilon_1\varepsilon_2\varepsilon_3| \\ &\leq 3|I|^3\varepsilon + 3|I|^3\varepsilon^2 + |I|^3\varepsilon^3 \leq 4|I|^3\varepsilon = 128(|I|^3/32)\varepsilon \\ &\leq 128\,\varepsilon \cdot \sup_I |h| \leq \frac{1}{2}\sup_I |h|, \end{aligned}$$

if $\varepsilon < 1/256$. Therefore $\sup_I |\widetilde{h}| \leq \sup_I |h| + \sup_I |h - \widetilde{h}| \leq (1 + 128\varepsilon)\sup_I |h|$, which is less than $\lambda \sup_I |h|$ if ε is small enough. Similarly, $\sup_I |\widetilde{h}| \geq \sup_I |h| - \sup_I |h - \widetilde{h}| \geq \frac{1}{\lambda}\sup_I |h|$ if ε is small enough, proving the lemma. ∎

8. Estimating the effect of interior perturbations

In this section, we start the proof of Theorem 4.1. We have to verify that when we perform a 3-point perturbation inside G_m^n, the integral of the Chebyshev polynomial over G_m^n changes by a factor proportional to the perturbation, and that the effect on this integral of the perturbations in other intervals G_k^n, $k \neq m$, is small by comparison. Lemma 8.1 below and Lemma 9.1 in the next section provide exactly these estimates. We end this section by verifying an earlier claim that the map defined by (4.2) is a homeomorphism.

Also recall that in Section 4 we introduced the notation $T_n(x, y)$ with $x \in [-1, 1]$ and $y = (y_1, \ldots, y_N) \in \mathbb{R}^N$ to denote the nth Chebyshev polynomial $T_n(x)$ after we perturbed the three interior roots in each G_k^n by a factor y_k. As before, let $\widetilde{y}_k \in \mathbb{R}^N$ be the vector equal to y_k in the kth coordinate and zero elsewhere. Thus $T_n(x, \widetilde{y}_k)$ corresponds to perturbing only the interior roots of G_k^n and leaving all others fixed. Recall from (4.1) that we defined

$$A_m^n(y) = \frac{1}{|G_m^n|}\int_{G_m^n} T_n(x, y)\,dx.$$

Lemma 8.1. *Suppose $t > 0$, $K \in \mathbb{N}$, $n \geq 2K$ and $K \leq m \leq n - K$. Suppose that we perturb the three interior roots of G_m^n by a factor y_m with $|y| \leq t$, and we that leave all other roots of T_n fixed. Let $T_n(x, \widetilde{y}_m)$ denote the new polynomial obtained in this way. Then $|A_k^n(\widetilde{y}_m)| \leq \frac{7}{6}|G_m^n|$, and $A_k^n(\widetilde{y}_m)$ is strictly monotone. Moreover, for $y_m > 0$ we have*

$$A_m^n(\widetilde{y}_m) \geq A_m^n(0) + \frac{21}{20}y_m = \frac{1}{|G_m^n|}\int_{G_m^n} T_n(x)\,dx + \frac{21}{20}y_m \tag{8.1}$$

if K is sufficiently large and ε is sufficiently small. For $y_m < 0$ negative, we similarly get

$$A_m^n(\tilde{y}_m) \le A_m^n(0) - \frac{21}{20}\, y_m = \frac{1}{|G_m^n|} \int_{G_m^n} T_n(x)\, dx - \frac{21}{20}\, y_m. \tag{8.2}$$

The choice of $7/6$ and $21/20$ is only for convenience, to make some later arithmetic work out. Our proof of Theorem 1.1 only requires that these estimates hold with some constants strictly larger than 1.

Proof of Lemma 8.1. The claim that $|A_k^n(\tilde{y}_m)| \le \frac{7}{6}|G_m^n|$ holds because $|T_n(x, \tilde{y}_m)| \le 7/6$ if ε is small enough, by Lemma 7.2.

The monotonicity is also easy to see. Indeed, we simply want to verify that the situation shown in Figures 10 and 12 is correct: as we increase the factor y_m of the perturbation in G_m^n, the perturbed polynomials change strictly monotonically, and hence the same is true for their integrals over G_m^n. However, any two such perturbed polynomials can be written as P_1P_0 and P_2P_0, where P_1 and P_2 are cubics corresponding to the three perturbed roots in G_m^n, and P_0 is the product over all other roots. Since P_0 has no zeros in the interior of G_m^n, it does not change sign there, and thus it suffices to show that P_1 and P_2 move monotonically as a function of the perturbation factor y_m. This was Lemma 6.2.

Next we prove the quantitative bound (8.1); the proof of (8.2) is identical, except for the obvious changes of sign. Recall that G_m^n is divided into four nodal sub-intervals, $I_{4m-3}^n, \dots, I_{4m}^n$. Because the rescaled version of G_m^n does not exactly match $[-2, 2]$, the estimates are easier for the two middle intervals I_{4m-2}^n and I_{4m-1}^n, so we deal with these first. We also suppose the perturbation is by a positive factor $y_m > 0$.

On I_{4m-1}^n, the rescaled version of Lemma 6.6 says that the positive node is multiplied by at least $1 + (4 - a^{-2})y_m + O(y_m^2)$ everywhere on the subinterval, so the integral increases by an additive factor of at least

$$[3 + O(\alpha) + O(y_m)]\, y_m \int_{I_{4m-1}^n} T_n(x)\, dx \ge [3 + O(\alpha) + O(y_m)]\, \frac{2y_m}{\pi}\, |I_{4m-1}^n|,$$

where we have used that $a = 1 + \alpha$ and that (by Lemma 3.2) the integral of T_n over a nodal interval I_k^n is at least $(2/\pi)|I_k^n|$. Similarly, by the rescaled version of Lemma 6.7, the absolute value of the negative node on I_{4m-2}^n becomes smaller, and its integral increases by a positive additive factor of size at least

$$[(4a^{-2} - 1)y_m + O(y_m^2)]\, \frac{2}{\pi}\, |I_{4m-2}^n| \ge [3 + O(\alpha) + O(y_m)]\, \frac{2y_m}{\pi}\, |I_{4m-2}^n|.$$

Next we deal with the outer sub-intervals, namely I_{4m-3}^n and I_{4m}^n. Let $J \subset I_{4m-3}^n$ denote the part of I_{4m-3}^n that lands in $[-2, a]$ when we rescale G_m^n as described above (the left root maps to a, the center root is mapped to 0, and the right root maps to 1). Possibly $I_{4m-3}^n \setminus J$ is empty if I_{4m-3}^n maps into $[-2, a]$. The size of this "leftover" interval is small if all four subintervals of G_m^n are about the same size, which happens if G_m^n is not too near -1 or $+1$ by Lemma 2.4. This is where we use the assumption that $K \le m \le n - K$ for some large K.

More precisely, for any $\eta > 0$, Lemma 2.4 says all the nodal intervals in G_m^n have the same length up to a multiplicative factor of $1 - \eta$, if K is large enough. Hence we may assume that they all have length at least $(1-\eta)|G_m^n|/4$. Thus the part of I_{4m-3}^n that is not in J has length at most $O(\eta|I_{4m-3}^n|)$, and it maps to an interval of length $O(\eta)$ (possibly empty) to the left of -2 under the rescaling. On this part of I_{4m-3}^n, Corollary 6.10 says that the perturbed polynomial is larger than T_n, so the integral over this segment increases under the perturbation, since T_n is positive on I_{4m-3}^n.

On J, we use the rescaled version of Corollary 6.4, which says the 3-point perturbation makes T_n larger on J by a multiplicative factor of $1+(\frac{1}{3}+O(\alpha)+O(y_m))y_m$. Thus the integral $\int_J T_n(x)\,dx$ will increase by an additive factor of

$$\begin{aligned}\left[\frac{1}{3}+O(\alpha)+O(y_m)\right] y_m \int_J T_n(x)\,dx &\geq \left[\frac{1}{3}+O(\alpha)+O(y_m)\right]\frac{2y_m}{\pi}\,|J| \\ &\geq \left[\frac{1}{3}+O(\alpha)+O(y_m)\right]\frac{2y_m}{\pi}\,(1-\eta)\,|I_{4m-3}^n|,\end{aligned}$$

where we have used Lemma 3.2 again to give a lower bound for area of a node in terms of the length of the base interval. Since we have already shown that the integral over $I_{4m-3}^n\setminus J$ changes in the same (positive) direction, the integral over all of I_{4m-3}^n changes by at least the bound given for J. By a very similar argument, the integral of the negative node in I_{4m}^n is made smaller (in absolute value), by at least an additive factor of the same size.

Thus the increase in the integral over all of G_m^n is at least

$$\begin{aligned}&\left[\frac{2}{3}+6+O(\alpha)+O(y_m)\right]\frac{2y_m}{\pi}\,(1-\eta)\,|G_m^n|/4 \\ &\qquad\geq \left(\frac{10}{3\pi}+O(\alpha)+O(y_m)\right)(1-\eta)\,y_m\,|G_m^n| \geq \frac{21}{20}\,y_m\,|G_m^n|,\end{aligned}$$

if η and y_m are small enough (note $10/(3\pi)\approx 1.06103 > 21/20 = 1.05$ and $|\alpha|\leq\eta$). ∎

Figure 12 illustrates a computation of the change in the integral in a special case, and indicates the estimate of $y_n|G_m^n|$ in the previous lemma is within a factor of three of being sharp.

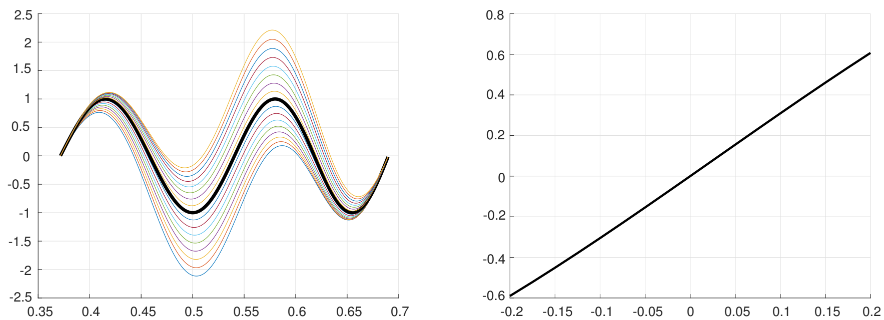

Figure 12. On the left are 3-point perturbations for $y=-.2,-.175,\ldots,.2$. The original Chebyshev polynomial is highlighted. On the right is a plot of $A_m^n(y)-A_m^n(0)$ for these perturbations. It has slope close to 3, while our proof shows the slope is ≥ 1. This example was taken with $n=33$ and $m=5$.

Corollary 8.2. *If $t > 0$ is small enough and $n, K \in \mathbb{N}$ are both large enough, then the map f as defined in equation (4.2) is a homeomorphism from Q_t^N to a cube Q' containing Q_t^N.*

Proof. By definition, the kth coordinate of f depends only on the kth coordinate of y, so the image is a cube, i.e., a product of compact intervals. By Lemma 8.1, every coordinate of f is a monotone function of the kth coordinate of y. Hence f is injective, and thus a homeomorphism. Finally, (8.1) and (8.2) imply the image under f_k of $[0, t]$ has length at least $21t/20$. Moreover, by Corollary 3.5, we know $A_m^n(0) < \eta/3$ if $n \geq 6/\sqrt{\eta}$. (Recall that $\eta > 0$ is our upper bound for the length ratio between nodal intervals inside G_m^n. This can be taken as close to zero as we wish by taking K large enough.) If $\eta \leq t/10$, and $n \geq 6/\sqrt{\eta}$, then $|A_m^n(0)| < t/20$. Therefore $f_k([0, t])$ contains $[A_m^n(0), t]$. The same argument shows $f_k([-t, 0]) \supset [-t, A_m^n(0)]$, and this implies $f(Q_t^N)$ contains Q_t^N. ∎

9. Estimating the effect of exterior perturbations

Next we see how perturbations of the roots outside G_m^n affect T_n inside this interval. This will complete the proof of Theorem 4.1.

Lemma 9.1. *Suppose $y \in Q_t^N$. If $t > 0$ is small enough and $K \in \mathbb{N}$ is large enough, then for $K < m \leq N - K$ we have*

$$|A_m^n(\tilde{y}_m) - A_m^n(y)| = \left|\frac{1}{|G_m^n|}\int_{G_m^n} T_n(x, \tilde{y}_m) - T_n(x, y)\, dx\right| \leq \frac{t}{2}.$$

Proof. For a fixed m, we want to estimate the contribution of all perturbations in G_k^n, for $k \neq m$, to the distortion function in G_m^n. Suppose $M \in \mathbb{N}$ (we will fix a value below). We divide the intervals $\{G_k^n\}_{k \neq m}$ into two groups according to whether $|k - m| \leq M$ or $|k - m| > M$. The second case (the more distant intervals) is easier, and we deal with it first.

Suppose I and J are the two intervals formed from G_k^n and G_m^n after rescaling the real line so that the center and right interior roots of G_k^n map to 0 and 1. Then I is approximately $[-2, 2]$ if $K < k \leq N - K$, by Corollary 2.4. With this normalization, Lemma 6.11 says that the distortion $R(x)$ at a point $x \in J$ of a perturbation by a factor t in I is at most

$$1 + \frac{2t + O(\alpha t) + O(t^2)}{(x - 1)x(x + 1)}. \tag{9.1}$$

Recall that here $\alpha = a - 1$, so $|\alpha|$ is as small as we wish by taking K large enough.

Therefore, the distortion bound on G_m^n for perturbations in G_k^n is largest at the endpoint of G_m^n closest to G_k^n, and is bounded by the formula above, except that x is replaced by the distance from G_m^n to the center root of G_k^n, divided by the distance between the center and right roots of G_k^n. We can simplify this a little by replacing the distance between G_m^n and the center of G_k^n by the distance from G_m^n to G_k^n. This is smaller, so gives a slightly larger bound. By Corollary 2.7, and because each G_j^n is made up of four nodal intervals, we have

$$\frac{|G_k^n|}{\operatorname{dist}(G_k^n, G_m^n)} \leq \frac{16}{4|k - m| - 1} \leq \frac{16}{4|k - m| - 4} = \frac{4}{|k - m| - 1}.$$

Recall from calculus that for $x > 0$, $1 + x = \exp(\log(1 + x)) \le \exp(x)$. Also, if f is decreasing on $[M, \infty)$, then $\sum_{j=M}^{\infty} f(j) \le \int_{M-1}^{\infty} f(x)\,dx$. By definition, the distortion functions for perturbing distinct sets of roots multiply to give the total distortion function, so we see that the total distortion on G_m^n due to perturbations in all G_k^n with $|k - m| > M$ is bounded by the product

$$\prod_{k:\,|k-m|>M} \left(1 + \frac{Ct}{(|k-m|-1)^3}\right) = \prod_{j \ge M} \left(1 + \frac{Ct}{j^3}\right)$$
$$\le \exp\Big(\sum_{j \ge M} \frac{Ct}{j^3}\Big) \le \exp\Big(Ct \int_{M-1}^{\infty} \frac{1}{x^3}\,dx\Big) \le \exp\Big(\frac{Ct}{2(M-1)^2}\Big).$$

The final term is less than $1 + t/M$ if M is large enough. Thus the distant intervals contribute almost no distortion.

Next we consider the distortion due to "nearby" intervals, i.e., the effect on G_m^n of perturbations in G_k^n with $|m - k| \le M$. This is more delicate than the "distant" intervals, and getting the first few terms (corresponding to intervals adjacent and nearly adjacent to G_m^n) to be small enough is one reason why we have used 3-point perturbations, instead of the simpler 2-point perturbations.

Fix $\eta > 0$. If K is large enough (depending on η and M), then by Lemma 2.4 we can assume all the nodal intervals contained in intervals G_k^n with $|m - k| \le M$ have lengths within a factor of $1 - \eta$ of each other. To simplify calculations, we normalize G_m^n as before, with the center and right-hand interior roots mapping to 0 and 1 respectively, and G_m^n maps to approximately $[-2, 2]$. Thus the renormalized nodal intervals have length approximately 1 (within a multiplicative factor of size $1 + \eta$). Since $1 - \eta \le (1 + \eta)^{-1}$, we have that $1 - \eta \le |I|/|J| \le 1 + \eta$ for any two nodal intervals $I, J \subset G_k^n$.

With these assumptions, if $|k - m| \le M$, then the distance between the center of G_m^n and a G_k^n is at least $x = 4(1 - \eta)|k - m|$. If $G_k^n = G_{m+j}^n$ is to the right of G_m^n, then G_{m+j}^n is approximately the interval $[4j - 2, 4j + 2]$ for $j = 1, 2, \ldots,$ M (with error at most η), and the maximum of our bound for the distortion on G_m^n by a perturbation in G_{m+j}^n occurs at the right endpoint of G_m^n, since this is the endpoint of G_m^n that is closest to G_{m+j}^n. Our distortion bound is smallest at the left endpoint of G_m^n, which is the furthest point of G_m^n from G_{m+j}^n. At this endpoint, our estimates say the distortion is at most

$$1 + \frac{[2 + O(\eta) + O(t)]t}{(4j-3)(4j-2)(4j-1)}.$$

The smallest size of our estimate occurs at the endpoint of G_m^n farthest from G_{m+j}^n and equals

$$1 + \frac{[2 + O(\eta) + O(t)]t}{(4j+1)(4j+2)(4j+3)} = 1 + \frac{[2 + O(\eta) + O(t)]t}{(4(j+1)-3)(4(j+1)-2)(4(j+1)-1)}.$$

Similar estimates hold for perturbations in intervals to the left of G_m^n, i.e., in G_{m-j}^n for $j = 1, 2, \ldots, M$.

Below, we will want to estimate the product of these terms over all the indices $k = m - M, \ldots, m + M$, except for $k = m$. We can get a slightly better estimate by pairing symmetrically placed terms of the form $m \pm j$, and we take advantage of this as follows. For the moment, we consider only the denominators in the bounds above.

It is easy to check from the explicit formula that the rational function giving the distortion bound due to perturbations in G^n_{m+j} is convex as a function of x on G^n_m (e.g., its partial fraction expansion is a sum of three convex terms on this interval). Similarly, the distortion bound for perturbations in G^n_{m-j} is convex on G^n_m. Therefore, the sum of these bounds is convex on this interval, and thus the sum takes its maximum value at one of the endpoints. See Figure 13.

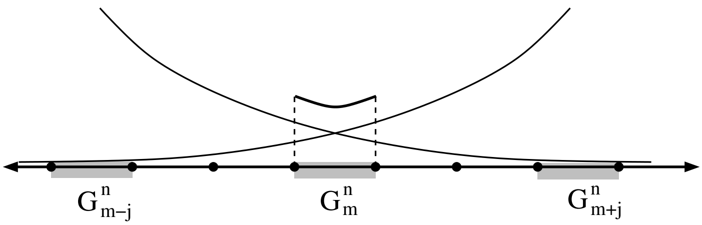

Figure 13. The sum of the distortion bounds corresponding to G^n_{m-j} and G^n_{m+j} is convex on G^n_m, and hence bounded on G^n_m by its values at the endpoints.

Using the elementary observation that

$$(1+x)(1+y) \leq \exp(\log((1+x)(1+y))) \leq \exp(x+y),$$

we can bound the product of the distortion bounds for the distortions in both G^n_{m+j} and G^n_{m-j} by

$$\exp\Big(\frac{1}{(4j-3)(4j-2)(4j-1)} + \frac{1}{(4(j+1)-3)(4(j+1)-2)(4(j+1)-1)}\Big).$$

Taking the product of distortions for all j is thus bounded by the exponential of the corresponding sum of these fractions over $j = 1, \dots, M$. Because the second fraction above is the same as the first, but with j replaced by $j+1$, each fraction is repeated twice, except for the first and last. Thus we get the upper bound for the sum

$$\frac{1}{1\cdot 2\cdot 3} + 2\sum_{j=2}^{M} \frac{1}{(4j+1)(4j+2)(4j+3)} \leq \frac{1}{6} + \frac{2}{210} + \frac{2}{990} + \frac{2}{2730} + \cdots$$

The sum becomes larger by replacing M by ∞, and we can bound the infinite sum (that clearly converges) by computing a finite number S of terms and bounding the remaining tail by the estimate

$$\sum_{j=S}^{\infty} \frac{1}{(4j+1)(4j+2)(4j+3)} \leq \sum_{j=S}^{\infty} \frac{1}{(4j+1)^3} \leq \sum_{j=4S+1}^{\infty} \frac{1}{j^3} \leq \int_{4S}^{\infty} x^{-3}\,dx \leq \frac{1}{32S^2}.$$

Taking $S = 100$ gives the upper bound $.1799 < 1/5$.

Using this (and the fact $1 + x \le \exp(x)$), the distortion bound on G_m^n due to perturbations within M steps of G_m^n is bounded by

$$\prod_{k:0<|k-m|\le M} \Big(1 + \frac{(2 + O(\eta) + O(t))t}{|(|k-m|+1)|k-m|(|k-m|-1)|}\Big)$$
$$\le \exp\Big(\sum_{k:0<|k-m|\le M} \frac{(2+O(\eta)+O(t))t}{|(|k-m|+1)|k-m|(|k-m|-1)|}\Big) \le \exp\Big(\frac{(2+O(\eta)+O(t))t}{5}\Big).$$

By taking η and t small enough, we can make this less than $1 + \lambda t$ for any $\lambda > 2/5$. We previously proved the distortion contributed by the distant intervals could be taken to be less than $1 + t/M$, so by taking M large enough (say $M \ge 10$), the total distortion from perturbations outside G_m^n is less than $1 + 3t/7$. Thus on G_m^n we have

$$\Big(1 - \frac{3t}{7}\Big) T_n(x, \tilde{y}_m) \le T_n(x, y) \le \Big(1 + \frac{3t}{7}\Big) T_n(x, \tilde{y}_m),$$

and hence, integrating over G_m^n,

$$\Big(1 - \frac{3t}{7}\Big) A_m^n(\tilde{y}_m) \le A_m^n(y) \le \Big(1 + \frac{3t}{7}\Big) A_m^n(\tilde{y}_m).$$

If we take t small enough so that $|A_m^n(\tilde{y}_m)| \le \frac{7}{6}|G_m^n|$, then this implies

$$|A_m^n(y) - A_m^n(\tilde{y}_m)| \le \frac{3t}{7}|A_m^n(\tilde{y}_m)| \le \frac{3t}{7}\frac{7}{6}|G_m^n| = \frac{t}{2}|G_m^n|,$$

as desired. ■

This completes the proof of Theorem 4.1. In the next section we use it to prove our main result, Theorem 1.1: polynomials with all critical points in a compact interval I are dense in $C_{\mathbb{R}}(I)$.

10. Applying Brouwer's fixed point theorem

Recall that $Q_t^n = [-t, t]^n$, that $\|(x_1, \dots, x_n)\| = \max_{1\le k\le n} |x_k|$ denotes the supremum norm on $\mathbb{R}^n$, and that $\|f - g\|_Q = \sup_{x\in Q} |f(x) - g(x)|$. By Brouwer's fixed point theorem [3], any continuous map of Q_t^n into itself has a fixed point. There are now various short proofs of this result, see, e.g., [16].

Lemma 10.1. *Suppose that $I_t = [-t, t]$, and for $k = 1, \dots, n$, that $J_k \subset \mathbb{R}$ is a compact interval that contains I_t. Let $Q = Q_t^n = \prod_{k=1}^N I_t$ and $Q' = \prod_{k=1}^N J_k$. Suppose $f = (f_1, \dots, f_N)$ is a homeomorphism from Q to Q', and that $g\colon Q \to \mathbb{R}^N$ is a continuous map such that $\|f - g\|_Q \le t/2$. Then $Q_{t/2}^n \subset g(Q)$.*

Proof. Suppose $a \in Q_{t/2}^n$. We want to show there is $\hat{x} \in Q$ so that $g(\hat{x}) = a$. For $x \in Q_t^n$, define $F(x) = f^{-1}(a + f(x) - g(x))$. By assumption, for $x \in Q$,

$$\|a + f(x) - g(x)\| \le \|a\| + \|f(x) - g(x)\| \le t/2 + t/2 = t,$$

so $a + f(x) - g(x) \in Q' = f(Q)$. Thus $F(x)$ is well defined and $F(Q) \subset Q$. Since f is a homeomorphism, f^{-1} is continuous, and hence F is continuous. Thus Brouwer's fixed point theorem implies F has a fixed point $\hat{x} \in Q$. At this point,

$$f^{-1}(a + f(\hat{x}) - g(\hat{x})) = \hat{x} \implies a + f(\hat{x}) - g(\hat{x}) = f(\hat{x}) \implies a = g(\hat{x}). \quad \blacksquare$$

We want to apply this to the functions f and g defined in equations (4.2) and (4.3). Lemma 8.1 showed that f is a homeomorphism of $Q = Q_t^N$ onto a cube Q' containing Q, at least if $t > 0$ is small enough. Theorem 4.1 shows that $\|f - g\|_Q \le t/2$. Thus we are in a position to apply Lemma 10.1 to these functions, in order to prove the following (slightly stronger) version of Theorem 1.1.

Corollary 10.2. *Any Lipschitz function F on $[-1, 1]$ can be uniformly approximated to within $1/n$ by a polynomial P of degree $O(n)$ with all its (real or complex) critical points in $[-1, 1]$. Moreover, P and F agree at both endpoints of every interval G_m^n, except for a uniformly bounded number at the beginning and end of $[-1, 1]$. If F is A-Lipschitz, then we can choose P to be CA-Lipschitz, with a constant C that is independent of F and n.*

Proof. We claim it suffices to prove that every t-Lipschitz function can be uniformly approximated, for some positive value of t. To see this, note that if h is A-Lipschitz, then $\tilde{h} = (t/A) \cdot h$ is t-Lipschitz, and if $\tilde{p}$ approximates $\tilde{h}$ to within $\varepsilon t/A$, then $p = (A/t) \cdot \tilde{p}$ approximates h to within ε. This proves the claim. Also note that p is $(2A/t)$-Lipschitz if $\tilde{p}$ is 2-Lipschitz, which will be the case below. Thus we can take $C = 2/t$ in the statement of Theorem 1.1.

Choose t so that Theorem 4.1 holds, i.e., so that $\|f - g\|_{Q_{2t}^N} \le t$. Suppose $G_k^n = [s, t]$ and set

$$a_k = \frac{\Delta(F, G_k^n)}{|G_k^n|} = \frac{F(t) - F(s)}{t - s}.$$

Then $a \in Q_t^N$, since F is t-Lipschitz. Then by Lemma 10.1, for any $n \in 8\mathbb{N} + 1$ sufficiently large, there is a $y \in Q_t^N$ so that the perturbed Chebyshev polynomial $T_n(x, y)$ satisfies

$$A_k^n(y) = \frac{1}{|G_k^n|} \int_{G_k^n} T_n(x, y)\, dx = a_k, \quad \text{for all } k = 1, \dots, N.$$

Thus the anti-derivative

$$P(x) = F(0) + \int_0^x T_n(t, y)\, dt$$

is a polynomial of degree n that has all its critical points in $[-1, 1]$ and satisfies $P(x) = F(x)$ at each endpoint of any G_k^n (except possibly for a bounded number K at each end of $[-1, 1]$).

By Lemma 7.2, $|P'(x)| = |T_n(x, y)|$ is bounded by 2 if t is small enough; thus P is 2-Lipschitz. Thus $|P(x) - f(x)| \le (2 + t)|G_k^n|/2$ on $|G_k^n|$, except for K intervals near each end, where we get the bound $K(2 + t) \max_k |G_k^n|$. Since $\max_k |G_k^n| \le 4\pi/n \to 0$ (Lemma 2.3), we see that P approximates F to within $O(1/n)$ on $[-1, 1]$. $\blacksquare$

This completes the proof of Theorem 1.1.

11. Weak-$*$ convergence: Proof of Theorem 1.2

In this section, we prove Theorem 1.2, i.e., that every bounded, measurable, real-valued function on $[-1,1]$ is the weak-$*$ limit of real polynomials with only real critical points. Recall that a sequence $\{f_n\} \subset L^\infty$ converges weak-$*$ to $f \in L^\infty$ if for every $g \in L^1$, $\int_I f_n g\,dx \to \int_I fg\,dx$. The definitions and results we quote below can be found in standard texts such as [10].

Proof of Theorem 1.2. It suffices to prove this for functions on $[-1,1]$. Fix a real-valued $f \in L^\infty([-1,1])$. We want to find polynomials $\{P_n\}$ that are uniformly bounded on $[-1,1]$ and so that $\int gP_n \to \int gf$ for any $g \in L^1([-1,1])$. Let $\{G_k^n\}$ be the partition of $[-1,1]$ into unions of four adjacent Chebyshev nodal intervals, as in the proof of Theorem 10.2. Using that theorem, there is a $K \in \mathbb{N}$ and a real-valued polynomial P_n so that $\|P_n\|_\infty \le C\|f\|_\infty$ and

$$\int_{G_k^n} P_n = \int_{G_k^n} f \tag{11.1}$$

for every $k = K, \ldots, (n/4) - K$. The union of the $2K$ intervals G_k^n where this estimate does not hold have total length tending to zero as n increases to ∞. Fix $g \in L^1([-1,1])$ and note that both $\int gf$ and $\int gP_n$ tend to zero over these intervals, so we can restrict attention to the union $I \subset [-1,1]$ of sub-intervals where (11.1) does hold.

For $M < \infty$, define g_M by $g_M = g$ on $\{x \in I : |g(x)| < M\}$ and $g_M = 0$ elsewhere. Since $|g - g_M| \le g \in L^1$ and $g_M \to g$ pointwise almost everywhere, the Lebesgue dominated convergence theorem implies $\|g - g_M\|_1 \to 0$ as $M \nearrow \infty$. So given any $\varepsilon > 0$, we can choose M so large that $\|g - g_M\|_1 \le \varepsilon$. Since $\|P_m\|_\infty \le C\|f\|_\infty$, we have

$$\int_I P_n g - \int_I fg = \int_I (P_n - f)g_M + \int_I (P_n - f)(g - g_M) = \int_I (P_n - f)g_M + O(\varepsilon\|f\|_\infty).$$

Therefore, it is enough to show $\int_I (P_n - f)g \to 0$ as $n \nearrow \infty$.

Let ψ_n be a step function approximation to g_M that is constant on the segments $\{G_k^n\}$ and so that $\|g_M - \psi_n\|_1 \to 0$ as $n \nearrow \infty$. Note that $\int_I (f - P_n)\psi_n = 0$, since the integral of $f - P_n$ is zero on each sub-interval G_k^n where ψ_n is constant. Therefore,

$$\begin{aligned}\left|\int_I (P_n - f)g_M\right| &= \left|\int_I P_n(g_M - \psi_n) + \int_I (P_n - f)\psi_n + \int_I (\psi_n - g_M)f\right| \\ &\le \|P_n\|_\infty \|g_M - \psi_n\|_1 + \|f\|_\infty \|\psi_n - g_M\|_1 \le (C+1)\|f\|_\infty \|g_M - \psi_n\|_1,\end{aligned}$$

and the last term tends to zero as n increases. Thus $\int_I P_n g \to \int_I fg$ for any $g \in L^1(I)$, and hence $P_n \to f$ weak-$*$.

The final step is to show that we cannot take $C = 1$ in Theorem 1.2. Suppose we could. Define $f = 1$ on $[0,1]$ and $f = 0$ on $[-1,0)$, and suppose that p_n are polynomials with only reals zeros, that $\|p_n\|_\infty \le 1$, and that p_n converges weak-$*$ to f. By weak convergence, $\int_0^1 p_n\,dx \to 1 = \int_0^1 f\,dx$, but for any $\varepsilon > 0$,

$$\begin{aligned}\int_0^1 p_n\,dx &\le |\{x \in [0,1] : p_n(x) > 1 - \varepsilon\}| + (1-\varepsilon)|\{x \in [0,1] : p_n(x) \le 1 - \varepsilon\}| \\ &= 1 - \varepsilon|\{x \in [0,1] : p_n(x) \le 1 - \varepsilon\}|.\end{aligned}$$

For $\varepsilon > 0$ fixed, the only way the right-hand side can tend to 1 is if

$$|\{x \in [0,1] : p(x) \le 1-\varepsilon\}| = |\{x \in [0,1] : |p(x) - f(x)| \ge \varepsilon\}| \to 0.$$

Thus p_n converges to f in measure on $[0,1]$. By the Clunie–Kuijlaars theorem discussed in the introduction (Corollary 1.3, [5]), f must be an entire function. But f is discontinuous, a contradiction. Therefore, $C = 1$ is impossible. ∎

Question 11.1. *What is the optimal value of C in Theorem 1.2?*

12. Divergence almost everywhere

In this section, we prove the claim from the introduction that the polynomial approximants we construct in the proof of Theorem 1.1 have derivatives that diverge pointwise almost everywhere.

Recall from Section 7 that the derivatives of each of our approximants are of the form $T_n(x,y)$, i.e., a perturbed version of the Chebyshev polynomial T_n. We want to show that $\frac{d}{dx}T_n(x,y)$ is large whenever $T_n(x,y)$ is small, so that the set were $T_n(x,y)$ is close to zero has small measure.

Lemma 12.1. *Suppose $T_n(x,y)$ is a perturbed Chebyshev function as in Lemma 7.2. If $\|y\| < t$ and t is small enough, then $|\{x \in [-1,1] : |T_n(x,y)| < t\}| = O(t)$.*

Proof. Suppose $r_k^n(y)$ is the kth root of the perturbed polynomial $T_n(x,y)$. Recall that this is a perturbation of r_k^n, the kth root of $T_n(x)$ (possibly r_k^n is equal to $r_k^n(y)$). The point r_k^n is the left endpoint of the kth nodal interval I_k^n of $T_n(x)$.

Write $T_n(x,y) = (x - r_k^n)S_k^n(x)$, i.e., S_k^n is a constant A times the product of terms $(x - r_j^n)$ over all the roots other than r_k^n. Let $x_k \in I_k^n$ and $x_{k+1} \in I_{k+1}^n$ be points where $|T_n(x,y)|$ is maximized in I_k^n and I_{k+1}^n, respectively. Then

$$|S_k^n(x_k)| = \frac{|T_n(x_k,y)|}{|x_k - r_k|} \ge \frac{1/2}{|I_k|} \tag{12.1}$$

and

$$|S_k^n(x_{k+1})| = \frac{|T_n(x_{k+1},y)|}{|x_{k+1} - r_k|} \ge \frac{1/2}{|I_{k+1}|}, \tag{12.2}$$

since by Lemma 7.2 the extreme values of $|T|$ are bigger than $1/2$ if t is small enough. In both cases, the lower bound is bigger than $1/(2|J_k|)$ where $J_k = I_k^n \cup I_{k+1}^n$. By Lemma 2.4, adjacent nodal intervals have comparable length, so $|J_k| \simeq |I_k^n|$.

Since

$$\log|S_k^n(x)| = \log|A| + \sum_{j:j\neq k} \log|x - r_j|,$$

and each individual term is concave down on $[x_k, x_{k+1}]$, we see that $\log|S_k^n|$ is concave down here as well. Hence $\log|S_k^n|$, and therefore $|S_k^n|$, attains its minimum over $[x_k, x_{k+1}]$

at one of the endpoints, i.e., at either x_k or x_{k+1}. By equations (12.1), (12.2), and the remarks following them, we deduce that $|S_k^n(x)| \geq C/|J_k|$ for all $x \in [x_k, x_{k+1}]$. Thus

$$\{x \in J_k : |T_n(x,y)| < \varepsilon\} = \{x \in J_k : |x - r_k| \cdot |S_k^n(x)| < \varepsilon\} \subset \left\{x \in J_k : \frac{C|x - r_k|}{|J_k|} < \varepsilon\right\},$$

and hence

$$|\{x \in J_k : |T_n(x,y)| < \varepsilon\}| \leq |\{x \in J_k : |x - r_k| < \varepsilon|J_k|/C\}| = \frac{2\varepsilon|J_k|}{C}.$$

Note that the intervals J_k cover each I_k^n twice, so $\sum |J_k| \leq 2\sum_k |I_k^n| \leq 4$. Therefore,

$$|\{x \in [-1,1] : |T_n(x,y)| < \varepsilon\}| \leq \frac{2\varepsilon}{C}\sum_k |J_k| \leq \frac{8\varepsilon}{C} = O(\varepsilon). \qquad \blacksquare$$

Corollary 12.2. *If f in Theorem 1.1 is Lipschitz, the sequence $\{p_n\}$ can be chosen so that $\{p_n'\}$ is uniformly bounded and so that any sub-sequence diverges pointwise Lebesgue almost everywhere on* $[-1,1]$.

Proof. As in the proof of Theorem 1.1, it suffices to consider f to be t-Lipschitz for some fixed $t > 0$, and we showed in Section 10 that such a function can be approximated by polynomials whose derivatives are functions of the form $T_n(x, y_n)$, with $|y_n| \leq t$. From the proof of Lemma 12.1, we can deduce that the sets

$$N_n = \{x \in [-1,1] : T_n(x,y_n) < -1/2\} \quad \text{and} \quad P_n = \{x \in [-1,1] : T_n(x,y_n) > 1/2\}$$

each have density bounded uniformly away from zero in any interval $[a,b] \subset [-1,1]$ if n is large enough, i.e.,

$$\liminf_{n\to\infty} \frac{|P_n \cap [a,b]|}{b-a} \geq c > 0.$$

The same inequality holds for N_n. The Lebesgue differentiation theorem (see, e.g., Theorem 3.21 in [10]) then implies that for every subset $E \subset [-1,1]$ of positive measure, there is an interval $J \subset [-1,1]$ so that $|E \cap J| > (1 - c/2)|J|$, e.g., take a small enough interval around a point of density of E. Therefore $|E \cap P_n \cap J|$ and $|E \cap N_n \cap J|$ both have measure larger than $(c/2)|J| > 0$ for all large enough n (depending on E). However, if E^+ is the set of x's where $\{T_{n_k}(x, y_{n_k})\}$ converges to a non-negative limit for some sequence $\{n_k\}$, then E^+ is disjoint from N_n for all sufficiently large n, hence E^+ must have Lebesgue measure zero. Similarly, if E^- is the set where the sequence converges to a non-positive limit, then E^- is disjoint from P_n for all large enough n, and hence E^- also has zero measure. Thus the sequence of derivatives diverges almost everywhere. $\blacksquare$

The Clunie–Kuijlaars theorem discussed in the introduction implies that if the sequence $\{p_n'\}$ does not converge uniformly on $[-1,1]$ to a Laguerre–Pólya function, then at almost every point of $[-1,1]$ it either diverges, converges to 0, or $|p_n'|$ converges to ∞. Corollary 12.2 shows the first option can occur, and [1] shows that approximants can be chosen so that $p_n'(x)$ converges to zero almost everywhere, or so that it tends to either $+\infty$ or $-\infty$ almost everywhere.

13. Theorem 1.1 fails for some subsets of $\mathbb{R}$

For $X \subset \mathbb{R}$, let $CP(X)$ denote all uniform limits of real polynomials on X with all critical points contained in X. Theorem 1.1 says that $CP(X) = C_{\mathbb{R}}(X)$ when X is an interval, but we will show that this fails for some disconnected subsets of $\mathbb{R}$. If X has only one or two points, the situation is trivial (non-constant linear functions have no critical points and can approximate every function on X), but even for three points the answer is not immediately obvious. If $X = \{-1, 0, 1\}$, then any polynomial p with critical points in X has a derivative of the form

$$C \int (x+1)^a x^b (x-1)^c \, dx,$$

for some triple (a, b, c) of non-negative integers. Thus

$$\frac{p(1) - p(0)}{p(0) - p(-1)} = \frac{\int_0^1 (x+1)^a x^b (x-1)^c \, dx}{\int_{-1}^0 (x+1)^a x^b (x-1)^c \, dx},$$

and this takes only countably many different values. Thus p restricted to X cannot equal every possible real-valued function on these three points. However, with some work, one can show that the set of possible ratios is dense in $\mathbb{R}$, and this implies $CP(X) = C_{\mathbb{R}}(X)$. Very briefly, consider $p(x) = (1-x)^a (1+x)^{n-a}$ with n large and $0 \le a \le n$. Normalize p so the maximum of p in $[-1, 1]$ is 1. This function has a single bump with max that slides from -1 to 1 as a goes from 0 to n (for an example, see Figure 14). The width of the bump is about $1/\sqrt{n}$, but the distance between the peaks for consecutive a's is about $1/n$. Thus only about $1/\sqrt{n}$ of the mass moves across 0 as we increment a near $n/2$, and careful estimates show that we can approximate any positive ratio that we want. However, for four points, the situation is different.

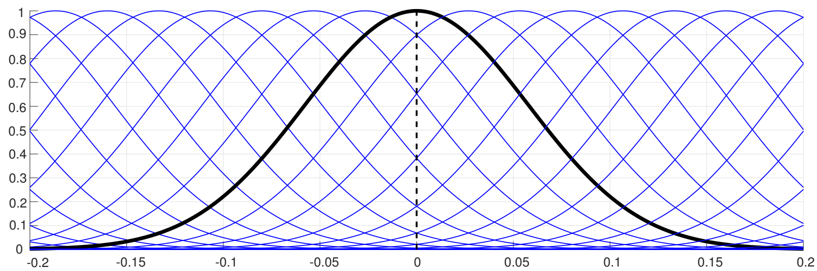

Figure 14. Plots of $C \int_0^t (x+1)^a (x-1)^{n-a} \, dx$, where $n = 300$, $a = 2, 4, \ldots, 298$ and $C = C(a, n)$ is chosen so the function's maximum is 1. The thicker graph corresponds to $a = n/2$. By taking n large and "sliding" a from 1 to n, the ratio of the areas over $[-1, 0]$ and $[0, 1]$ can approximate any positive value we wish.

Lemma 13.1. *Suppose that* $0 < \varepsilon < 1/4$. *Suppose that* X *is a compact set contained in* $[-1-\varepsilon, -1] \cup [1, 1+\varepsilon]$, *and that it contains at least two distinct points in each of these intervals. Then there is a real-valued function continuous on* X *that cannot be uniformly approximated by polynomials that have all their critical points inside* X.

Proof. Let $X_{-1} = X \cap [-1-\varepsilon, 1]$ and $X_1 = X \cap [1, 1+\varepsilon]$. Let $s = \inf X_{-1}$, $t = \sup X_{-1}$, $u = \inf X_1$, and $v = \sup X_1$. We claim there is a $\lambda < 1$ depending only on ε so that

$$\min(|p(s) - p(t)|, |p(u) - p(v)|) \leq \lambda |p(t) - p(u)| \tag{13.1}$$

holds for every real polynomial with all critical points in X. If such a λ exists, then a function f on X such that

$$|f(s) - f(t)| = |f(t) - f(u)| = |f(u) - f(v)| > 0$$

cannot be uniformly approximated by polynomials in $CP(X)$.

Take a p as above and let p' be its derivative. Rescaling by a constant, we may assume that p' is monic and that it has n zeros in X_{-1} and m zeros in X_1. Note that p' has a single sign in $[t, u]$; without loss of generality, we may assume $p' > 0$ here. For $x \in [0, \varepsilon]$, the distance of x to X_{-1} is ≥ 1 and its distance to X_1 is $\geq 1 - \varepsilon$. Thus on $[0, \varepsilon]$ we have $|p'| \geq (1-\varepsilon)^m$. Similarly, $|p'| \geq (1-\varepsilon)^n$ on $[-\varepsilon, 0]$. Thus

$$|p(u) - p(t)| \geq \left| \int_t^u p' \right| \geq \max\left(\left| \int_{-\varepsilon}^0 p' \right|, \left| \int_0^\varepsilon p' \right| \right) \geq \max\left(\varepsilon(1-\varepsilon)^n, \varepsilon(1-\varepsilon)^m \right).$$

Every point of $[1, 1+\varepsilon]$ is within distance ε of the m roots contained in X_1 and is within distance $2 + 2\varepsilon$ the n roots in X_{-1}. Therefore on X_1 we have $|p'| \leq \varepsilon^m (2+2\varepsilon)^n$. Similarly, on X_{-1} we have $|p'| \leq \varepsilon^n (2+2\varepsilon)^m$. Thus

$$|p(s) - p(t)| \leq \left| \int_s^t p' \right| \leq \varepsilon^{n+1} (2+2\varepsilon)^m,$$

$$|p(u) - p(v)| \leq \left| \int_u^v p' \right| \leq \varepsilon^{m+1} (2+2\varepsilon)^n.$$

The inequality $\min(x, y) \leq \lambda \max(w, z)$ follows from $xy \leq \lambda^2 wz$, so to prove (13.1) it suffices to verify that

$$\left(\varepsilon^{n+1}(2+2\varepsilon)^m\right)\left(\varepsilon^{m+1}(2+2\varepsilon)^n\right) \leq \lambda^2 \varepsilon^2 (1-\varepsilon)^{n+m}$$

i.e.,

$$\left(\frac{\varepsilon(2+2\varepsilon)}{1-\varepsilon} \right)^{n+m} \leq \lambda^2.$$

This holds for some $\lambda < 1$ if $\varepsilon(2+2\varepsilon) < 1 - \varepsilon$. Using the quadratic formula, we get

$$\frac{-3-\sqrt{17}}{4} < \varepsilon < \frac{-3+\sqrt{17}}{4} \approx .2808.$$

The right side is larger than $1/4$, so this proves the lemma. ∎

Question 13.2. *For which compact sets $X \subset \mathbb{R}$ is $CP(X) = C_{\mathbb{R}}(X)$? Does this fail for all disconnected sets X with more than three points?*

Acknowledgments. I thank the anonymous referee for a meticulous reading of the manuscript and for many very helpful suggestions that improved the clarity and correctness of the exposition.

References

[1] Bishop, C. J. and Bishop, D. L.: Approximation by singular polynomials. Preprint 2023.

[2] Bishop, C. J. and Lazebnik, K.: A geometric approach to polynomial and rational approximation. *Int. Math. Res. Not. IMRN* (2024), no. 12, 9936–9961. MR 4761785

[3] Brouwer, L. E. J.: Über Abbildung von Mannigfaltigkeiten. *Math. Ann.* **71** (1911), no. 1, 97–115. Zbl 42.0417.01 MR 1511644

[4] Brown, J. E. and Powell, V. F.: A result on real polynomials with real critical points. *J. Anal. Appl.* **5** (2007), no. 1, 41–52. Zbl 1159.30005 MR 2297082

[5] Clunie, J. G. and Kuijlaars, A. B. J.: Approximation by polynomials with restricted zeros. *J. Approx. Theory* **79** (1994), no. 1, 109–124. Zbl 0814.41008 MR 1294323

[6] Csordas, G. and Dimitrov, D. K.: Conjectures and theorems in the theory of entire functions. *Numerical Algorithms* **25** (2000), no. 1-4, 109–122. Zbl 0968.30012 MR 1827149

[7] Epstein, A. L.: Symmetric rigidity for real polynomials with real critical points. In *Complex manifolds and hyperbolic geometry (Guanajuato, 2001)*, pp. 107–114. Contemp. Math. 311, American Mathematical Society, Providence, RI, 2002. Zbl 1017.37018 MR 1940166

[8] Erdös, P. and Grünwald, T.: On polynomials with only real roots. *Ann. of Math. (2)* **40** (1939), no. 3, 537–548. Zbl 65.0056.03 MR 0000007

[9] Eremenko, A. and Gabrielov, A.: Rational functions with real critical points and the B. and M. Shapiro conjecture in real enumerative geometry. *Ann. of Math. (2)* **155** (2002), no. 1, 105–129. Zbl 0997.14015 MR 1888795

[10] Folland, G. B.: *Real analysis*. Second edition. Pure and Applied Mathematics (New York), Wiley, New York, 1999. Zbl 0924.28001 MR 1681462

[11] Hale, N. and Trefethen, L. N.: Chebfun and numerical quadrature. *Sci. China Math.* **55** (2012), no. 9, 1749–1760. Zbl 1257.41018 MR 2960859

[12] Hinkkanen, A. and Kayumov, I.: On critical values of polynomials with real critical points. *Constr. Approx.* **32** (2010), no. 2, 385–392. Zbl 1210.30002 MR 2677885

[13] Korevaar, J. and Loewner, C.: Approximation on an arc by polynomials with restricted zeros. *Indag. Math.* **26** (1964), 121–128. Zbl 0128.29803 MR 0161984

[14] Kozlovski, O., Shen, W. and van Strien, S.: Rigidity for real polynomials. *Ann. of Math. (2)* **165** (2007), no. 3, 749–841. Zbl 1129.37020 MR 2335796

[15] Marshall, D. E.: *Complex analysis*. Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 2019. Zbl 1429.30001 MR 4321146

[16] Milnor, J.: Analytic proofs of the "hairy ball theorem" and the Brouwer fixed-point theorem. *Amer. Math. Monthly* **85** (1978), no. 7, 521–524. Zbl 0386.55001 MR 0505523

[17] Pólya, G.: Über Annäherung durch Polynome mit lauter reelen Wurzeln. *Rend. Circ. Mat. Palermo* **36** (1913), 279–295. Zbl 44.0474.01

[18] Ritt, J. F.: Permutable rational functions. *Trans. Amer. Math. Soc.* **25** (1923), no. 3, 399–448. Zbl 49.0712.02 MR 1501252

[19] Titchmarsh, E. C.: *The theory of the Riemann zeta-function*. Second edition. The Clarendon Press, Oxford University Press, New York, 1986. Zbl 0601.10026 MR 0882550

[20] Trefethen, L. N.: *Approximation theory and approximation practice*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013. Zbl 1264.41001 MR 3012510

[21] Weierstrass, K.: Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen. *Sitzungsber. Akad. Berlin* (1885), 633–639 and 789–806. Zbl 17.0384.02

Received August 25, 2023; revised January 16, 2024.

David L. Bishop
Department of Economics, Yale University
PO Box 208268, New Haven, CT 06520, USA;
david.bishop@yale.edu, contactdavidbishop@gmail.com