Advances in Mathematics 490 (2026) 110819

Contents lists available at ScienceDirect

Advances in Mathematics

journal homepage: www.elsevier.com/locate/aim

Approximation by singular polynomial sequences

Christopher J. Bishop [a,*,1], David L. Bishop [b]

[a] *Mathematics Department, Stony Brook University, Stony Brook, NY 11794-3651, United States of America*
[b] *Economics Department, Yale University, New Haven, CT 06520-8268, United States of America*

A R T I C L E   I N F O

Article history:
Received 18 January 2026
Accepted 21 January 2026
Available online xxxx
Communicated by C. Fefferman

MSC:
primary 30C10, 30C62, 30E10
secondary 41A20

Keywords:
Weierstrass approximation
Critical points
Chebyshev polynomials
Uniform approximation
Laguerre-Pólya class

A B S T R A C T

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 with only real critical points and whose derivatives converge to zero almost everywhere on I. Alternatively, the approximants may be chosen so that the derivatives converge to plus infinity almost everywhere, or so that these behaviors each occur almost everywhere on specified sets. This extends work by the second author, showing that the derivatives can also be taken to diverge pointwise almost everywhere. Together, these results prove that a 1994 theorem of Clunie and Kuijlaars is sharp.

1. Introduction

This paper is a sequel to [1]. In that paper, the classic Weierstrass approximation theorem [14] was strengthened by proving the following.

* Corresponding author.
E-mail addresses: bishop@math.stonybrook.edu (C.J. Bishop), david.bishop@aya.yale.edu (D.L. Bishop).

1 CJB is partially supported by NSF Grant DMS 2303987.

https://doi.org/10.1016/j.aim.2026.110819
0001-8708/

Proposition 1.1. *Any real-valued, continuous function f on a compact interval $I \subset \mathbb{R}$, can be uniformly approximated by real polynomials $\{p_n\}$ so that all their critical points lie in I. If, in addition, f is K-Lipschitz, then we can take the $\{p_n\}$ to be $O(K)$-Lipschitz. Moreover, p_n' converges weak-$*$ to f' as elements of $L^\infty(I)$, but p_n' diverges pointwise almost everywhere.*

In this paper, we give a different variation of this result in which p_n' may be chosen to converge almost everywhere to either 0 or ∞. A theorem of Clunie and Kuijlaars [4] states that if $\{p_n'\}$ has only real zeros and converges pointwise to finite, non-zero, real values on a set $E \subset \mathbb{R}$ of positive Lebesgue measure, then $\{p_n'\}$ must actually converge uniformly on every compact subset of $\mathbb{C}$ to an entire function in the Laguerre-Pólya class (defined below). Thus if $\{p_n\}$ are polynomials with only real critical points that converge uniformly to a general function f (not the anti-derivative of a Laguerre-Pólya function), then at almost every point $x \in I$, the sequence $\{p_n'(x)\}$ either diverges or it converges to either 0, $-\infty$ or $+\infty$. The approximating sequences $\{p_n\}$ constructed in [1] have derivatives that diverge almost everywhere on I, showing the first alternative can occur. In this paper, we construct approximating sequences so that $\{p_n'\}$ converges to 0, or $-\infty$, or $+\infty$. Thus all the behaviors allowed by the Clunie-Kuijlaars theorem actually occur.

In analogy to singular functions in real analysis (non-constant, continuous functions that have derivative zero almost everywhere), we shall say that $\{p_n\}$ is a singular sequence of polynomials if $\{p_n\}$ converges uniformly to a continuous, non-constant function f, but $\{p_n'\}$ converges to zero almost everywhere. Our main result is that every real-valued, continuous function can be uniformly approximated by such a sequence.

Theorem 1.2. *If f is real-valued and continuous on I, then there is a sequence of polynomials $\{p_n\}$ with only real critical points, so that $p_n \to f$ uniformly on I and $\{p_n'\}$ converges to zero almost everywhere. If f is increasing on I, then the elements of $\{p_n\}$ may be chosen to be increasing on I as well.*

Increasing polynomials p_n obviously satisfy $p_n' \geq 0$, but it turns out that one cannot always take strict inequality in the final part of Theorem 1.2.

Theorem 1.3. *Suppose f is real-valued and continuous on I, and that $\{p_n\}$ are real-valued polynomials that converge uniformly to f, and that the polynomials $\{p_n\}$ only have real critical points. If f is not the anti-derivative of the restriction of a Laguerre-Pólya entire function to I, and if $J \subset I$ is a non-trivial interval on which f is non-constant, then J contains a critical point of p_n for all sufficiently large n (depending on J). In other words, the critical points of $\{p_n\}$ accumulate everywhere that f is non-constant.*

The theorem of Clunie and Kuijlaars also allows for the possibility that $\{p_n'\}$ converges pointwise almost everywhere to $-\infty$ or $+\infty$. We will show this can occur.

Theorem 1.4. *If f is real-valued and continuous on I, then there is a sequence of polynomials $\{p_n\}$ with only real critical points so that $\{p_n\}$ converges uniformly to f on I, and $\{p_n'\}$ converges to $+\infty$ almost everywhere on I.*

Without the restriction on the critical points, it is easy to obtain the weaker condition $|p_n'| \to \infty$ almost everywhere: if $q_n \to f$ uniformly, then it is not hard to verify that $p_n = q_n + \frac{1}{n}T_{k_n} \to f$ uniformly as well, and that $|p_n'| \to \infty$ almost everywhere if $k_n \nearrow \infty$ quickly enough, depending on the choice of $\{q_n\}$. Here, T_n is the nth Chebyshev polynomial, defined later in this introduction. Thus the point of Theorem 1.4 is to get the "one sided" divergence to $+\infty$, while restricting the critical points to the interval I. By first approximating $-f$ by a sequence $\{p_n\}$ with derivatives tending to $+\infty$ almost everywhere, and then changing signs, it is clear that Theorem 1.4 also holds with $+\infty$ replaced by $-\infty$. In Section 5, we will also note that similar constructions give sequences $\{p_n\}$ with only real critical points so that $p_n \to f$ uniformly and with p_n' tending to 0, $-\infty$ or $+\infty$ respectively, almost everywhere on any three disjoint, measurable sets whose union is I.

Polynomials with only real critical points have played a role in several problems, e.g., density of hyperbolicity in dynamics [10], rigidity of conjugate polynomials [7], Smale's conjecture on solving polynomial systems [8], and Sendov's conjecture on locations of critical points [3]. In holomorphic dynamics, the orbits of critical points play an essential role. Various constructions in the field make use of approximation theorems such as Weierstrass's and Runge's theorems, and it is desirable to control the locations of the critical points of the approximating functions. In [2], a version of Runge's theorem is proven where all critical points may be taken to lie within any open ϵ-neighborhood of a connected set K. In [1], this is further improved to $\epsilon = 0$ when $K = I \subset \mathbb{R}$ is an interval, i.e., Weierstrass's theorem holds even if we require all critical points to lie in I. However, [1] also constructs disconnected sets $K \subset \mathbb{R}$ where $C_{\mathbb{R}}(K)$ (real valued, continuous functions on K) is not the uniform closure of polynomials with all critical points in K. Classifying the sets K when this does occur remains an open problem.

The Laguerre-Pólya class, mentioned above, 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 [13]: 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

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

with $m \in \{0, 1, 2, \dots\}$, $a, b \in \mathbb{R}$ and $c \le 0$. 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.

A theorem of Korevaar and Loewner [9], extending earlier work of Laguerre [11] and Pólya [13], says that if $\{p_n\}$ are polynomials with only real zeros that converge uniformly

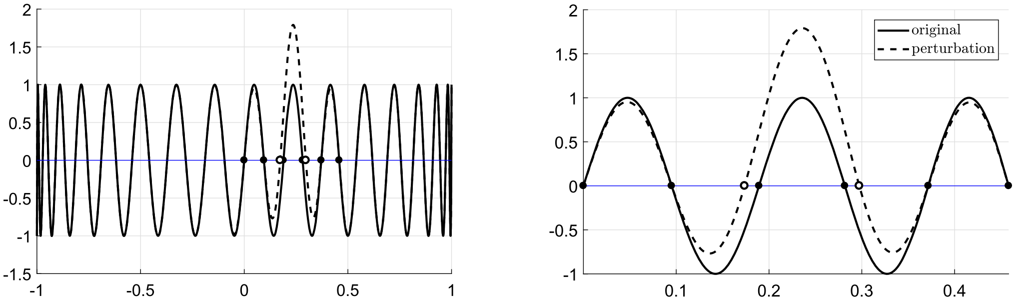

Fig. 1. A 2-point perturbation of T_{33}, shown over the full interval $[-1, 1]$ (left) and an enlargement around the perturbed roots (right). Separating two adjacent roots creates a larger node between them, while having little effect on the size of more distant nodes.

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). See also [5]. Clunie and Kuijlaars [4] 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. Since almost everywhere convergence on a set of finite measure implies convergence in measure, we obtain the pointwise version of their theorem quoted earlier.

If a real polynomial p of degree $n+1$ has all n critical points in $[-1, 1]$, then its derivative can be written in the form

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

where $C \in \mathbb{R}$ and $\{z_k^n\}_{k=1}^n \subset [-1, 1]$. The polynomials used in this paper are all of this form, where $\{z_k^n\}$ are perturbations of the roots $\{r_k^n\}$ of nth Chebyshev polynomial T_n. We briefly recall the definition of these polynomials.

Let $J(z) = \frac{1}{2}(z+1/z)$ be the Joukowsky map. It is easy to verify that this map sends a point $z = x+iy$ on the unit circle to $x \in [-1, 1]$, and that 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 $J^{-1} : U \to \mathbb{D}^*$. Therefore $T_n = J((J^{-1})^n)$ is a n-to-1 holomorphic map of U to U that is continuous across $\partial U = [-1, 1]$, so by Morera's theorem (e.g., Theorem 4.19 of [12]) it is holomorphic on the whole plane, and hence it is a degree n polynomial. Unwinding the definition, we see that T_n maps $[-1, 1]$ into itself and is given by $T_n(x) = \cos(n \arccos x)$. It takes the 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 has its zeros at $\{r_n^k\} = \{\cos(\pi \frac{2k-1}{2n})\}_{k=1}^n$ (the vertical projections of the midpoints on $\mathbb{T}$ between the roots of unity). See Fig. 1 for an example. This figure (and many others in this paper) was drawn using the `MATLAB` program `Chebfun` by L.N. Trefethen and his collaborators. See [6].

Fix a large positive integer n and consider the Chebyshev polynomial T_n. Order the n roots of T_n from left to right, and for $k = 1, \ldots, n-1$, let I_k^n denote the interval between the kth and $(k+1)$st roots of T_n. We call these the "nodal intervals", and call

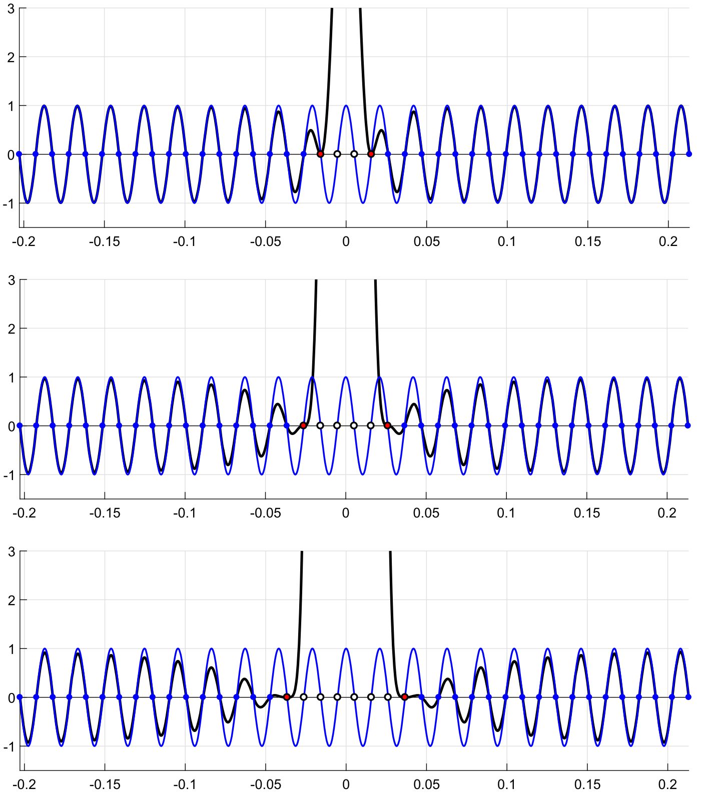

Fig. 2. A Chebyshev polynomial of degree 300 near the origin, and the polynomials obtained by moving N pairs of roots for $N = 1, 2, 3$. The white dots represent roots that are moved; all others are kept fixed. The thinner curve is the original Chebyshev polynomial and the thicker is the perturbed polynomial. The height of the new nodes is better illustrated in logarithmic coordinates in Fig. 3.

the restriction of T_n to 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 of T_n at the endpoints of I_k^n farther apart (and leave all the other roots of T_n fixed), then the node between them becomes higher, and the two adjacent negative nodes each get smaller (less negative). More distant nodes are changed slightly, but the effect diminishes with distance from I_k^n. This will be made precise in Section 3. See Fig. 1 for the basic idea.

This is the fundamental operation that we use to create the polynomials we want: choose an interval J bounded by two roots of T_n and move each root by equal amounts away from each other. This procedure was introduced in [1], where roots were moved using small perturbations. Here "small" means that a root r_k^n of T_n is only moved within the interval $[r_{k-1}^n, r_{k+1}^n]$, i.e., it is moved no further than the nearest adjacent root on either side, and usually it is only moved a small fraction of this distance. When the perturbations are small in this sense, then the perturbed Chebyshev polynomials created are uniformly bounded. This was important in [1] in order to prove that a K-Lipschitz function f can be approximated by polynomials (with only real critical points) that are

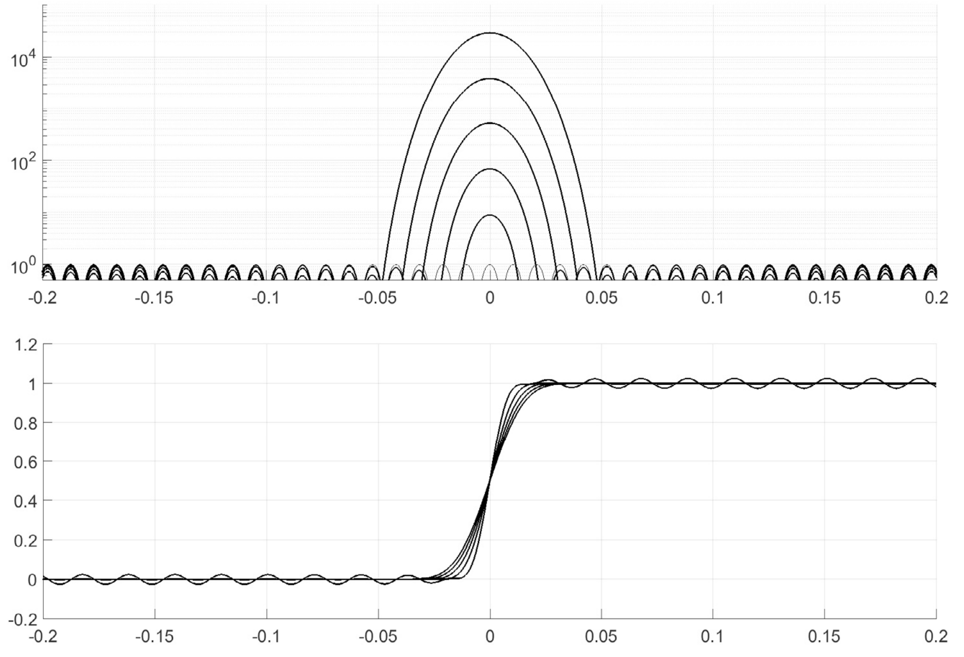

Fig. 3. On top is $\log |p|$ for perturbations of a degree 300 Chebyshev polynomial after moving N pairs of points, for $N = 1, \dots, 5$. The maximums are growing exponentially with N. On the bottom are the anti-derivatives of the perturbations, renormalized so $\int_{-1}^{1} p(t)dt = 1$. These anti-derivatives converge to a step function with jump at 0. In both pictures, the horizontal axis is restricted to $[-.2, .2]$.

$(C\dot{K})$-Lipschitz, for some fixed $C < \infty$. [1] shows we must have $C > 1$, but the optimal value remains unknown.

In this paper, we will be concerned with "large" perturbations, i.e., roots that are moved farther than the closest adjacent roots. We will choose an even number of adjacent Chebyshev roots $\{r^n_{k+1}, r^n_{k+2}, \dots, r^n_{k+2N}\}$, and then move half these to points close to (but larger than) $a = r^n_k$, and move the other half to points close to (but less than) $b = r^n_{k+2N+1}$. This creates a very large node inside the interval (a, b); we will show that the height of this node grows exponentially with N, e.g., Equation (3.6). By choosing the size and sign of these nodes correctly, and rescaling appropriately, the anti-derivatives form polynomial sequences satisfying Theorems 1.2 and 1.4. Some examples of multi-point perturbations are shown in Fig. 2.

Fig. 3 gives essentially the same plots (superimposed on top of each other) but with a logarithmic scale on the vertical axis. The heights of the new nodes do appear to grow exponentially in N (linearly on the logarithmic scale) and we will verify this in Section 3. The bottom picture in Fig. 3 shows anti-derivatives of the perturbed polynomials, normalized to have total integral 1. Because the un-normalized mass grows exponentially with N, the normalized functions are exponentially smaller than the originals. In particular, since the un-normalized perturbations are bounded by 1 outside of the interval I where the perturbations occur, the normalized polynomials are exponentially small

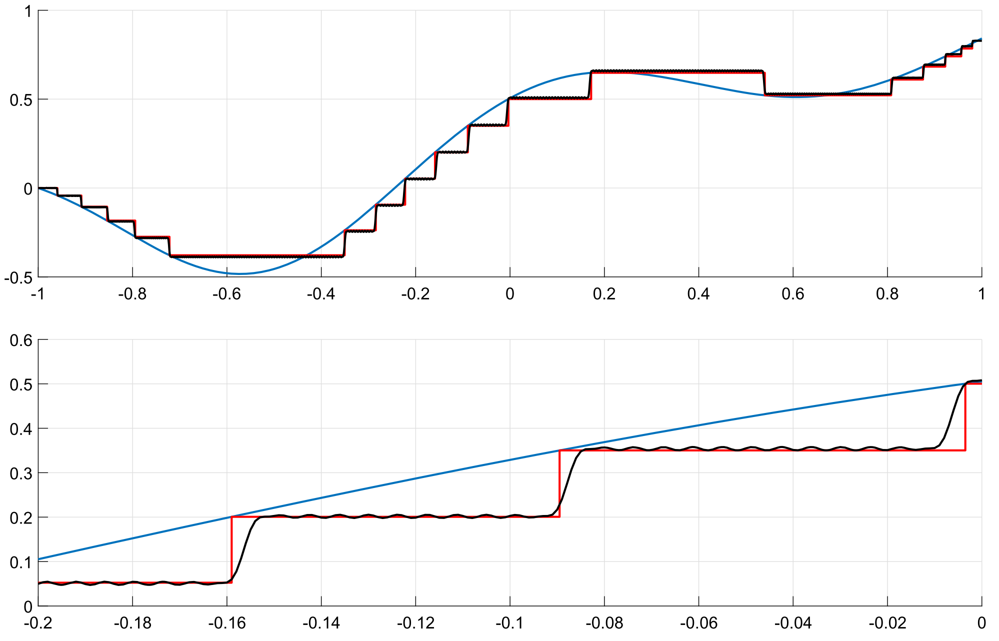

Fig. 4. At top, a smooth function is approximated by step function, which in turn is approximated by the anti-derivative p of a perturbed Chebyshev polynomial. In this example, we have moved four roots near each jump. The step function and p are easier to distinguish in an enlargement over $[-.2, 0]$ (bottom picture).

outside this interval. Fig. 3 suggests that we can approximate a step function using anti-derivatives of perturbed Chebyshev polynomials as described above. This will be made precise in later sections.

The idea behind Theorem 1.2 is that we can create a perturbed Chebyshev polynomial that has nodes with exponentially large area near specified points of $[-1, 1]$, and these nodes can be chosen to be either positive or negative. By multiplying by a scalar, we can make these nodes have area $\pm\epsilon$, while the function is much smaller away from these nodes. The integral of such a function looks like a step function with jumps of size $\pm\epsilon$, and by choosing the signs and areas of the large nodes correctly we can uniformly approximate any continuous function by polynomials of this form, i.e., a polynomial with only real critical points and with derivative less than ϵ except on a set of length ϵ. Taking a sequence of such polynomials with $\sum \epsilon_n < \infty$, and applying the Borel-Cantelli Lemma, gives a singular sequence of polynomials converging uniformly to f, proving the first part of Theorem 1.2 (once we have verified several details). See Fig. 4 for an example. Note the numerous critical points of the approximating polynomial; these are consistent with Theorem 1.3.

Our other results are proved using variations on this construction. For example, to obtain monotone approximations in Theorem 1.2, we follow the procedure above, but applied to T_n^2 and perturbing each double root as a single point. Then every root of the new polynomial has even degree, and hence the corresponding anti-derivative is monotone. This will allow us to approximate any monotone function f by a monotone singular sequence. For details, see Section 4.

If 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 A/B is bounded independent of the parameter. The notation $A \lesssim B$ means the same as $A = O(B)$. The more precise notation $A = O_C(B)$ will mean $|A| \le C|B|$. For example $x = 1+O_2(\frac{1}{n})$ is simply a more concise way of writing $1 - \frac{2}{n} \le x \le 1 + \frac{2}{n}$. The notation $A = \Omega_C(B)$ (or $A \gtrsim B$) means $A \ge CB$ or, equivalently, $B = O_C(A)$. We use $A \simeq B$ to mean that both $A \lesssim B$ and $A \gtrsim B$ hold, i.e., that A and B are comparable up to a fixed multiplicative constant, independent of the implicit parameter. In general, the notation $A = B$ means that two previously defined quantities are equal, and $A := B$ defines A in terms of B. This paper is mostly self-contained, except for a few standard estimates involving Chebyshev polynomials, quoted from [1].

We thank two anonymous referees for detailed reports that caught several minor errors and greatly improved the exposition.

2. Forced accumulation of real critical points

In this section, we will prove Theorem 1.3, but we start by gathering together various facts that we will need for the proof. Recall, from the introduction, the theorem of Clunie and Kuijlaars: if $\{q_n\}$ has only real roots and converges pointwise to finite, non-zero limits on a set of positive Lebesgue measure, then it must converge uniformly on all compact planar sets to a Laguerre-Pólya function. As a consequence of this, we will deduce the following result.

Lemma 2.1. *Suppose $J = [a, b] \subset \mathbb{R}$ is a compact interval and $\{q_n\}$ is sequence of real polynomials with only real roots, and that all the roots of all the q_n are in $\mathbb{R} \setminus J$. Suppose also that $m \le q_n \le M$ on J, for some $0 < m < M < \infty$ independent of n. Then there is a subsequence of $\{q_n\}$ that converges uniformly on compact subsets of the plane to a Laguerre-Pólya function.*

Proof. Suppose $q(x) = C \prod(x - r_k)$ is a polynomial with roots $\{r_k\} \subset \mathbb{R} \setminus J$. Since q does not change sign on J, without loss of generality we may assume $q > 0$ on J. Then

$$\log q(x) = \log |q(x)| = \log |C| + \sum \log |x - r_k|,$$

is a finite sum of continuous, concave down functions on J. This means that there is a $c_n \in [a, b]$ so that $\log q_n$ (and hence q_n) is increasing on $[a, c_n]$ and decreasing on $[c_n, b]$ (possibly $c_n = a$ or $c_n = b$). For every n, c_n is either $\le (a + b)/2$ or $\ge (a + b)/2$. Assume $\ge (a + b)/2$ occurs infinitely often. Then by passing to a subsequence we may assume every q_n is increasing on $J' = [a, \frac{1}{2}(a + b)]$, the left half of J. The other case, where we have that $\{q_n\}$ is decreasing on $J'' = [\frac{1}{2}(a + b), b]$, is almost identical to what follows, and is left to the reader.

Let $\mathbb{Q} \subset \mathbb{R}$ denote the rational numbers. For each $x \in J' \cap \mathbb{Q}$ we can take a subsequence so that $\{q_{n_k}(x)\}$ converges to a limit in $[m, M]$. By a diagonalization argument, we can find a subsequence that converges for every rational number in J', and at the endpoints of J'. The limiting function q must be increasing on $J' \cap \mathbb{Q}$, and it can be extended to an increasing function on all of J' by the formula

$$q(x) = \inf\{q(y) : y \in J' \cap \mathbb{Q} \cap [x, \infty)\}.$$

An increasing function has only countably many discontinuities (all jump discontinuities), so q is continuous almost everywhere on J', and $0 < m \le q \le M < \infty$.

If q is continuous at x, then we claim $q_n(x)$ converges to $q(x)$. To prove this, suppose $\epsilon > 0$ and use the continuity of q at x to choose $\delta > 0$ so that $|x - y| < \delta$ implies $|q(x) - q(y)| < \epsilon$. For $y \in \mathbb{Q} \cap (x, x + \delta)$, we have (since q_n is increasing)

$$q_n(x) \le q_n(y) \le q(y) + \epsilon \le q(x) + 2\epsilon$$

for large enough n (depending on ϵ), and hence $\limsup q_n(x) \le q(x)$. Similarly, if $z \in \mathbb{Q} \cap (x - \delta, x)$, then

$$q_n(x) \ge q_n(z) \ge q(z) - \epsilon \ge q(x) - 2\epsilon$$

for large enough n, and hence $\liminf q_n(x) \ge q(x)$. Thus $q_n(x) \to q(x)$ at every point of continuity of q. Since q is continuous on a set of positive measure, the conclusion of the lemma follows from the theorem of Clunie and Kuijlaars. □

We will say a real-valued, continuous function f on an interval $J = [a, b]$ is "inflection type" if there is a division point $c \in [a, b]$ so that f is convex up on $[a, c]$ and concave down on $[c, b]$ (there may be many such points if f is linear on some subinterval of J). We allow $c = a$ or $c = b$, hence convex and concave functions on J are also considered inflection type.

Lemma 2.2. *If $\{f_n\}_1^\infty$ are all inflection type on $J = [a, b]$, and $f_n \to f$ uniformly on J, then f is also inflection type.*

Proof. Let $\{c_n\}$ be a division point for f_n. By taking a subsequence, if necessary, we may assume $c_n \to c \in [a, b]$. First assume $a < c < b$. Then for any $\epsilon \in (0, c - a)$, f_n is convex up on $[a, c - \epsilon]$, if n is large enough so that $c_n > c - \epsilon$. Uniformly limits of convex functions are convex, so we deduce f is convex up on $[a, c]$. A similar argument shows f is concave down on $[c, b]$. If $c \in \{a, b\}$ then one of these arguments shows f is convex up or concave down on all of J, hence it is still of inflection type. □

An increasing function on an interval need not be strictly increasing on any subinterval, (e.g., the Cantor singular function), but an increasing, inflection-type function does have this property.

Lemma 2.3. *Suppose f is inflection type and increasing on $J = [a,b]$. Then there is $[x,y] \subset [a,b]$ so that f is constant on both $[a,x]$ and $[y,b]$, and so that f is strictly increasing on $[x,y]$. In particular, either f is constant on J (if $x = y$) or it is strictly increasing on some non-trivial sub-interval J (if $x < y$).*

Proof. If not, then there is a non-trivial interval $[s,t] \subset [a,b]$ such that f is constant on $[s,t]$ but non-constant on both $[a,s]$ and $[t,b]$. Thus $f(s) > f(a)$, and this implies that f is not convex up on $[a,u]$ for any $s < u < t$. Thus the division point for f satisfies $c \le s$. Therefore f must be concave down on all of $[s,b]$. Since it is constant on $[s,t]$ and increasing on $[a,b]$ this implies it is constant on $[s,b]$, a contradiction. This proves the lemma. □

Proof of Theorem 1.3. Suppose f is real-valued and continuous on $[-1,1]$, that $\{p_n\}$ are real-valued polynomials converging uniformly to f on $[-1,1]$, and that these polynomials have only real critical points. Assume $J \subset [-1,1]$ is a non-trivial interval, that f is not constant on J, and that all the critical points of every p_n are contained in $\mathbb{R} \setminus J$. Then p_n' is non-zero in J and by multiplying f by -1 (if necessary) and passing to a subsequence, we may assume every p_n' is positive in J. By Lemma 2.1 it suffices to show that there is a non-trivial subinterval of J where $\{p_n'\}$ is uniformly bounded above and uniformly bounded away from zero.

As in the proof of Lemma 2.1, J divides into left and right sub-intervals so that p_n' is increasing on the first sub-interval and decreasing on the second (possibly, just one of these sub-intervals occur). Thus each p_n is inflection type on J. Thus by Lemma 2.2, f is also inflection type on J. Since we assume f is not constant on J, Lemma 2.3 implies there is a non-trivial subinterval $J' = [a,b] \subset J$ where f is strictly increasing. Let $J'' = [c,d] = [\frac{2}{3}a + \frac{1}{3}b, \frac{1}{3}a + \frac{2}{3}b]$ be the middle third of J'. We will show that $\{p_n'\}$ has the desired lower and upper bounds on this interval.

First we prove the lower bound. Let $\epsilon = \min(f(c)-f(a), f(d)-f(c), f(b)-f(d))$. Since f is strictly increasing on $[a,b]$ this is positive. Assume n is so large than $|f - p_n| \le \epsilon/4$ on $[a,b]$. Suppose $s \in [c,d]$. If p_n' is increasing on $[a,s]$, then

$$(s-a)p_n'(s) \ge \int_a^s p_n' = p_n(s) - p_n(a) \ge f(s) - f(a) - \frac{\epsilon}{2} \ge f(c) - f(a) - \frac{\epsilon}{2} \ge \frac{\epsilon}{2}.$$

Hence

$$p_n'(s) \ge \frac{\epsilon/2}{s-a} \ge \frac{\epsilon}{2(c-a)} = \frac{3\epsilon}{2(b-a)} > \frac{\epsilon}{b-a}.$$

Otherwise, if p_n' is decreasing on $[s,d]$, we have

$$(b-s)p_n'(s) \ge \int_s^b p_n' = p_n(b) - p_n(s) \ge f(b) - f(s) - \frac{\epsilon}{2} \ge f(b) - f(d) - \frac{\epsilon}{2} \ge \frac{\epsilon}{2},$$

and hence

$$p_n'(s) \geq \frac{\epsilon/2}{b-s} \geq \frac{\epsilon}{2(b-d)} = \frac{3\epsilon}{2(b-a)} > \frac{\epsilon}{b-a}.$$

Since p_n' increases then decreases over J' (possible just increasing or just decreasing), one of these two options must hold, so we get the lower bound $p_n'(s) \geq m := \epsilon/(b-a)$.

To prove the upper bound, we use the fact that $\log p_n'$ is concave down on J. Suppose M_n is the maximum of p_n' over J'' and that the maximum is attained at $x \in J''$. Then $\log p_n'$ is bounded between $\log m$ and $\log M_n$, and by concavity the graph of $\log p_n'$ lies above the triangle with vertices $(c, \log m)$, $(x, \log M_n)$ and $(d, \log m)$. Hence $\log p_n' \geq \frac{1}{2}(\log M_n + \log m)$ on an interval $I \subset J''$ of length $|I| = \frac{1}{2}|J''| = (d-c)/2$. Therefore $p_n' \geq \sqrt{mM_n}$ on I, which implies

$$\int_I p_n' \geq |I|\sqrt{mM_n}.$$

By the Fundamental Theorem of Calculus, this integral equals $p_n(d) - p_n(c)$. Thus if n is so large that $|f(d) - p_n(d)| \leq \epsilon/4$ and $|f(c) - p_n(c)| \leq \epsilon/4$, then

$$\begin{aligned} M_n \leq \frac{(p_n(d)-p_n(c))^2}{m|I|^2} &\leq \frac{(f(d)-f(c)+\epsilon/2)^2}{m|I|^2} \\ &\leq 4\frac{(f(b)-f(a))^2}{m((d-c))^2} \\ &\leq 36\frac{((f(b)-f(a))^2}{m(b-a)^2} \\ &= 36\frac{(f(b)-f(a))^2}{\epsilon(b-a)}. \end{aligned}$$

This proves $M_n = \sup_{[c,d]} p_n'$ is bounded independent of n, as desired. By Lemma 2.1, this implies that f must be a Laguerre-Pólya function, contrary to our assumption. This contradiction proves Theorem 1.3. $\square$

Example. If $f(x) = \int e^{x^2}\,dx$ then $\log f' = x^2$ is not concave down in any sub-interval of $[-1,1]$. Although f is entire, if $p_n \to f$ uniformly on $[-1,1]$, and if every p_n has only real critical values, then these critical values must accumulate everywhere on $[-1,1]$, even though f itself only has a critical point at zero.

3. 2-point and multi-point distortions

In this section, we record some simple algebra that shows how a polynomial changes as we move some of its roots. This will verify certain claims made in the introduction.

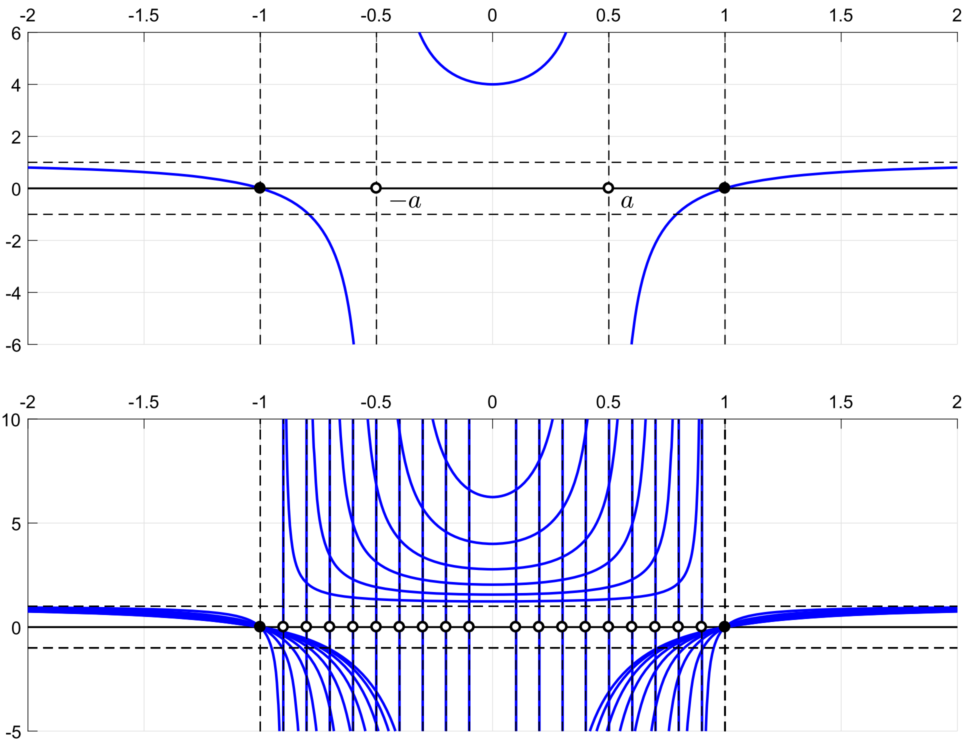

Fig. 5. On top is a plot of $R(t) = (x^2 - 1)/(x^2 - a^2)$ for $a = .5$ and below it are superimposed plots for $a = .1, .2, \ldots, .9$. The horizontal dashed lines are at heights ± 1. Outside the interval $[-1, 1]$ all these graphs are between $1 - x^{-2}$ and 1.

If a polynomial p has zeros at $\pm a$, for $a \in (0, 1)$, and if we move these roots respectively to ± 1, then we obtain a new polynomial $\widetilde{p} = R \cdot p$, where R is the rational function

$$R(x) = \frac{(x-1)(x+1)}{(x-a)(x+a)} = \frac{x^2-1}{x^2-a^2} = \frac{x^2-a^2+a^2-1}{x^2-a^2} = 1 - \frac{1-a^2}{x^2-a^2}. \qquad (3.1)$$

It is also easy to check that R is even, and that on the interval $(-a, a)$ we have $R(x) \geq R(0) = a^{-2}$. See Fig. 5.

Lemma 3.1. *For* $|x| \geq 1$ *we have* $1 - x^{-2} \leq R(x) \leq 1$.

Proof. To prove the lower bound, note that

$$\begin{aligned} |x| > 1 &\Rightarrow -a^2x^2 \leq -a^2 \\ &\Rightarrow x^2(1-a^2) \leq x^2 - a^2 \\ &\Rightarrow \frac{1-a^2}{x^2-a^2} \leq \frac{1}{x^2} \\ &\Rightarrow R(x) \geq 1 - \frac{1}{x^2}. \end{aligned}$$

For the upper bound, observe that if $|x| \geq a$, then $x^2 - a^2 > 0$, so $R(x) \leq 1$. $\square$

Lemma 3.2. *If $\alpha := 1 - a \in (0,1)$, then*

$$|R(x)| < 1 \textit{ for } |x| > (3+a)/4 = 1 - \alpha/4. \tag{3.2}$$

Proof. As noted above, if $|x| \geq a$, then $R(x) \leq 1$. Thus $R(x) < 1$ if $|x| > a = 1 - \alpha$. So we only need to check that $R(x) > -1$ for $|x| > 1 - \alpha/4$. To prove this, note that if $x^2 > a^2$, then $R(x) > -1$ is equivalent to

$$\begin{aligned} & x^2 - 1 > a^2 - x^2 \\ \Leftrightarrow\ & x^2 > \tfrac{1}{2}(a^2+1) \\ \Leftrightarrow\ & |x| > \sqrt{\tfrac{1}{2}((1-\alpha)^2+1)} \\ \Leftrightarrow\ & |x| > \sqrt{1-\alpha+\alpha^2/2}. \end{aligned}$$

The right side is less than $1 - \alpha/4$ if and only if

$$\begin{aligned} & \sqrt{1-\alpha+\alpha^2/2} < 1 - \alpha/4 \\ \Leftrightarrow\ & 1-\alpha+\alpha^2/2 < 1-\alpha/2+\alpha^2/16 \\ \Leftrightarrow\ & \alpha^2(\tfrac{1}{2} - \tfrac{1}{16}) < \alpha/2 \\ \Leftrightarrow\ & \alpha < \tfrac{8}{7}, \end{aligned}$$

which is certainly true, since we assumed $\alpha = 1 - a \in (0,1)$. Thus $|x| > 1 - \alpha/4$ implies $R(x) > -1$. $\square$

We can apply a linear transformation to the points in the preceding estimates. Note that translating the points $\{\pm 1, \pm a\}$, all by the same amount just translates R. Similarly, dilating to get new points $\{\pm\lambda, \pm\lambda a\}$ just replaces $R(x)$ by $R(lx/\lambda)$. In particular, if $b_1 < a_1 < a_2 < b_2$ are chosen so that $c = \frac{1}{2}(a_1 + a_2) = \frac{1}{2}(b_1 + b_2)$, and $a_1, a_2 = c \pm r$, $b_1, b_2 = c \pm s$, then these four points are images of $\{\pm 1, \pm a\}$ (where $a = r/s$) under a linear map. Hence, the distortion function only changes by pre-composition with a linear map, and we deduce the following result.

Corollary 3.3. *With notation as above, the distortion function R corresponding moving $a_1, a_2 = c \pm r$ to $b_1, b_2 = c \pm s$ satisfies*

(1) $R(x) \geq (s/r)^2$ *on* (a_1, a_2),
(2) $0 < R(x) < 1$ *on* $\mathbb{R} \setminus [b_1, b_2]$,
(3) $|R(x)| < 1$ *on* $\{|x - c| > (1 - \alpha/4)s\}$ *where* $\alpha = 1 - s/r$.
(4) $1 - x^{-2} \leq R(x) < 1$ *on* $\{|x - c| \geq s\}$.

If we move multiple pairs of zeros in this way, then the distortion function for the combined moves is just the product of the distortion functions for each pair. More precisely, suppose N is a positive integer and we have $2N$ points $\{a_k\}_{|k|=1}^N$ (note that there is no point a_0) such that

$$-1 < a_{-N} < a_{-N+1} < \cdots < a_{-1} < 0 < a_1 < \cdots < a_N < 1.$$

Suppose we move the pair (a_{-k}, a_k) to a pair $(b_{-k}, b_k) \subset [-1, 1]$, again with the property that $b_{-k} + b_k = a_{-k} + a_k$. Given a small $\delta > 0$, we will also assume that $1 - \delta \leq |b_k| \leq 1$, so that the new zeros are all quite close to ± 1. In particular, $|b_k - b_{-k}| \geq 2 - 2\delta$. We can place b_{-k} and b_k so that this happens as long as $|a_{-k} + a_k| < \delta$; this will occur in our construction. Indeed, we will take the $\{a_k\}$ to be approximately evenly spread in $[-1, 1]$, i.e., $a_k \approx \text{sign}(k) \cdot (2|k| - 1)/(2N + 1)$ for $|k| = 1, \ldots, N$ and we will choose $\delta < 1/N$.

For the moment, we make the weaker assumption that

$$1/(2N + 2) \leq |a_k| \leq |k|/N. \tag{3.3}$$

This implies $|a_k - a_{-k}| \leq 2k/N$. In later applications, we will renormalize the points in an interval $J = [a, b]$, and use the analogous condition

$$1/(2N + 2) \leq \frac{|a_k - (a + b)/2|}{(b - a)/2} \leq |k|/N. \tag{3.4}$$

The $2N$-point distortion function R_N, resulting from moving the points $\{a_k\} \subset [-1, 1]$ to points $\{b_k\} \subset [-1, 1]$, is the product of N different 2-point distortion functions as described above, one for moving each pair $\{a_{-k}, a_k\}$. Thus we have $0 < R_N(x) < 1$ on $\{|x| > 1\}$. Moreover, taking logarithms and using some calculus, it is easy to check that for $|x| > 1$

$$\max(0, 1 - N/x^2) \leq (1 - x^{-2})^N \leq R_N(x) \leq 1. \tag{3.5}$$

We also claim that our perturbed polynomial $\widetilde{p} = p \cdot R_N$ satisfies $|\widetilde{p}| \gg |p|$ near the orgin. Assume that we have chosen δ so that $\delta < 1/N$. By Part (1) of Lemma 3.3, on the interval $|x| < 1/(2N + 2)$, the $2N$-point distortion satisfies

$$\begin{aligned} R_N(x) &\geq \prod_{k=1}^N \left(\frac{b_k - b_{-k}}{a_k - a_{-k}}\right)^2 \geq \frac{(2 - 2\delta)^{2N}}{\prod_{k=1}^N (2k/N)^2} \\ &= (2 - 2\delta)^{2N} N^{2N} 2^{-2N} (N!)^{-2} \\ &> (1 - \frac{1}{N})^{2N} N^{2N} (N!)^{-2} \end{aligned}$$

for $x \in [a_{-1}, a_1]$. Using the upper bound in Stirling's approximation for $N!$,

$$\sqrt{2\pi N}\left(\frac{N}{e}\right)^N \le N! \le \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \exp(\frac{1}{12N})$$

and the inequality $e^{-1/6N} \ge e^{-1/6} > 1/2$, the above lower bound for R_N becomes

$$R_N(x) \ge \frac{(1-\frac{1}{N})^{2N} e^{2N-1/6N} N^{2N}}{2\pi N N^{2N}} \ge \frac{(1-\frac{1}{N})^{2N} e^{2N}}{4\pi N}$$

Recalling from calculus that $1/4 \le (1-1/N)^N \nearrow e^{-1}$ for $N \ge 2$, this becomes

$$R_N(x) \ge \frac{[(1-1/N)^N]^2 e^{2N}}{2\pi N} \ge \frac{e^{2N}}{32\pi N}. \tag{3.6}$$

For large N, this is bigger than e^N, so the size of the perturbed node grows at least exponentially with N.

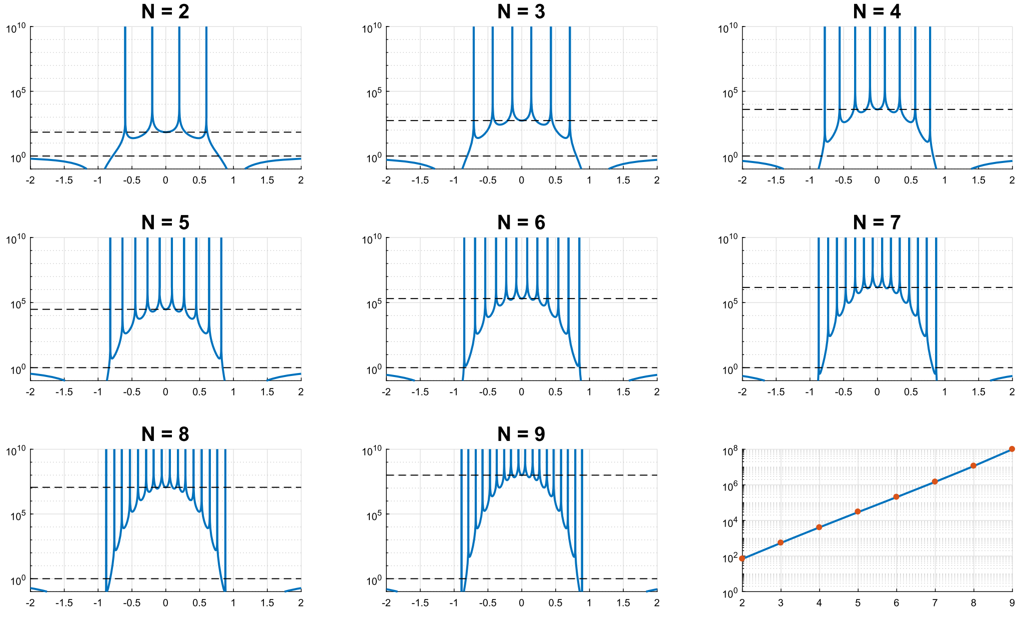

Fig. 6. Logarithmic plots of $2N$-point distortion functions R_N for $N = 1, \dots, 9$. Here we are moving the points $\{\pm(2k-1)/(2N+1)\}_{k=1}^N$ to $\{\pm 1\}$; these points are evenly spaced in $[-1, 1]$, and moved to the nearest endpoint. In each plot, one dashed line is at height 1, and the other shows the minimum value of R_N over the central interval. The final plot shows a linear approximation (with slope ≈ 2) to the log-plot of these minima versus N, indicating the minima grow like $\approx \exp(2N)$.

Thus the central node of the $2N$-perturbed polynomial is exponentially larger than the original node. The area of the original node is comparable to $|a_1 - a_{-1}|$, so the area of the new node is larger than this by a factor of at least $e^{2N}/32\pi N$. A logarithmic plot of the actual distortion in the cases $M = 1, \dots, 9$ is shown in Fig. 6. As expected, the growth of the distortion function near the origin is exponential. Numerically, the growth rate appears to be $\approx e^{2N}$, which is what we expect from (3.6).

4. Approximation by polynomials with small derivatives

The idea of the proof of Theorem 1.2 is to approximate a continuous function f by a step function g, and then approximate g by the anti-derivative p of a scalar multiple of a polynomial q that is constructed by perturbing the zeros of a Chebyshev polynomial T_n. The basic idea was illustrated in Fig. 4.

This proof is the most intricate in the paper, so to make the argument easier to follow, we break it into a number of steps. We list them here, and give the details later in this section. After translation and rescaling, it suffices to prove the theorem for the interval $I = [-1, 1]$.

(1) Approximate f to within ϵ by a step function g, where the jumps of g are all of size $\pm\epsilon$. Let K denote the number of jumps of g and let $-1 < s_1 < \cdots < s_K < 1$ denoted their locations. Define $s_0 = -1$ and $s_{K+1} = 1$. Let $\delta > 0$ be the minimum distance between the points of $\{s_j\}_0^{K+1}$.

(2) We will define K disjoint intervals $\{G_j\}_1^K \subset [-1, 1]$, so that for each $j = 1, \ldots, K$, the interval G_j contains (and is approximately centered at) the jump point s_j. Each G_j will be a union of $2N + 1$ nodal intervals of a Chebyshev polynomial T_n where $N := 4\lceil \log n \rceil$. The disjointness will follow if n is sufficiently large, depending on δ. Since each G_j is a union of an odd number of nodal intervals of T_n, there is a central nodal interval which we denote G_j^C. We will choose G_j so the sign of T_n on the central interval G_j^C is the same as the sign of the jump of g at s_j. The leftmost and rightmost nodal intervals of T_n contained in G_j will be denoted G_j^L and G_j^R respectively.

(3) We let η_j denote the length of the shortest nodal interval contained in G_j. Clearly $\eta_j \le |G_j|/(2N + 1)$, and will show that if n is large enough, then the ratio $|G_j|/\eta_j(2N + 1)$ is as close to 1 as we wish.

(4) For each j, we choose points $\{b_k\}_{|k|=1}^N \subset G_j$ to be the new roots of the perturbation of T_n. Half of these b_k's will be located in a subinterval $J_- \subset G_j^L$ of length $\eta_j/10$, and the other half within an interval $J_+ \subset G_j^R$ of equal length.

(5) If we perturb only the roots of T_n in G_j, then the perturbed polynomial will have one large node covering most of G_j. We will estimate the area of this node, showing it is exponentially large in N.

(6) By continuously moving the new roots $\{b_k\}$ back towards the center of G_j, the area under this large central node decreases continuously. Thus we can make it attain any value we want within a specified range. In particular, we will be able to attain the value $\pm\epsilon \cdot e^{N/2}$, where ϵ and N are as chosen in Steps 1 and 2 above. The nodal interval corresponding to this large central node will be denoted $J_j \subset G_j$.

(7) We create a polynomial q from T_n by making the perturbations described in Step 6 in every G_j simultaneously. Then set $p' = e^{-N/2} \cdot q$. We will show that $|q| \le 1$ on $X = [-1, 1] \setminus \cup_j J_j$, and hence that $|p'|$ is exponentially small there.
(8) Next we will show that the perturbations performed in one interval G_k have only a small effect on the size of the large central node in any other G_j, $j \ne k$. Thus each of the "large nodes" of p' has integral $\int_{J_j} p' \approx \pm\epsilon$ with errors that tend to zero as n increases. This completes the proof that the step function g (and hence the original function f) is uniformly approximated by $p = \int p'$, a polynomial with only real critical points.
(9) We verified in Step 7 that $|p'|$ is very small except possibly on the set $\cup_j J_j$, which has small length. By taking an appropriate sequence of such approximants and applying the Borel-Cantelli lemma, we will deduce that f can be approximated by a singular sequence of polynomials.
(10) The final step is to verify that if f is increasing, then it can be approximated by a singular sequence of increasing polynomials. This requires only a minor modification of the proof sketched above, obtained by repeating the proof, but now applied to T_n^2, and moving roots in pairs. The resulting polynomial q will then have only roots of multiplicity two, so we can choose q, and hence p', to be non-negative everywhere.

Before filling in the details of the preceding sketch, we recall some estimates concerning nodal intervals and integrals for Chebyshev polynomials. These are quoted from [1], but are standard facts. Recall that the nodal intervals of T_n are the $n-1$ intervals between adjacent roots of T_n, and are denoted $\{I_k^n\}_{k=1}^{n-1}$ from left to right. The intervals are symmetric with respect to zero, so the estimates below only have to be given for nodal intervals hitting $[-1, 0]$.

Lemma 4.1 *(Lemma 2.3, [1]). For* $1 \le k \le (n-1)/2$, $\frac{4k}{n^2} \le |I_k^n| \le \frac{k\pi^2}{n^2}$.

Lemma 4.2 *(Lemma 2.4, [1]). If* $1 \le k \le k + j \le n/2$ *then*

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

Lemma 4.3 *(Lemma 3.2, [1]).* $\int_{I_k^n} |T_n| \ge \frac{2}{\pi}|I_k^n|$.

Proof of Theorem 1.2.
• **Step 1:** Without loss of generality, we may assume $f(-1) = 0$. Fix $\epsilon \in (0, 1)$. Choose an ordered set of points $\{-1 = s_0 < s_1 < \cdots < s_K < s_{K+1} = 1\}$ so that

$$|f(s_{j+1}) - f(s_j)| = \epsilon, \text{ for } j = 0, 1, \ldots, K-1$$
$$|f(s_K) - f(s_{K-1})| \le \epsilon,$$

and

$$|f(t) - f(s_j)| < \epsilon, \text{ for } t \in [s_j, s_{j+1}).$$

Define a step function $g(t)$ on $[-1, 1]$ by $g(t) = f(s_j)$ for $t \in [s_j, s_{j+1})$, for $j = 0, \dots K$. Clearly $\|f - g\|_I \leq \epsilon$. Let $\delta = \min_{1 \leq j \leq K+1}(s_j - s_{j-1})$.

• **Step 2:** Suppose n is a large positive integer, that will be fixed at the end of the proof, depending only on ϵ, δ and K from Step 1. Consider the Chebyshev polynomial T_n. Choose K nodal intervals $\{I^n_{k_j}\}_{j=1}^K$ so that $I^n_{k_j}$ either contains s_j or it is adjacent to a nodal interval that does contain s_j. We choose the interval so that T_n has the same sign on $I^n_{k_j}$ as the sign of the jump of g at s_j.

Set $N = 4\lceil \log n \rceil$, and let G_j be the union of $I^n_{k_j}$ and the N nodal intervals on either side of it. Thus G_j is the union of $2N + 1$ nodal intervals; the central interval $I^n_{k_j}$ is denoted G^C_j for brevity, and the leftmost and rightmost are denoted G^L_j and G^R_j respectively. By Lemma 4.1, the intervals $\{G_j\}$ are pairwise disjoint if n is large enough (depending only on δ). Indeed, this lemma implies the length of G_j is less than $4\pi^2\lceil \log n \rceil/n$, so the distance between any two of these intervals is at least $\delta/2$, if n is large enough. (To simplify notation, we have omitted a superscript n, writing G_j instead of G^n_j. This convention will also apply to other points and intervals below, but the implicit dependence on n should be clear.)

• **Step 3:** For each $j = 1, \dots, K$, we are going to move the $2N$ roots of T_n inside G_j to new points near the endpoints of G_j. This procedure was described in Section 1 and illustrated in Fig. 2. The endpoints of G_j will be left fixed. The $2N$ roots of T_n in the interior of G_j are denoted (again omitting the dependence on j and n from the notation)

$$a_{-N} < \cdots < a_{-1} < a_1 < \cdots < a_N$$

as in Section 3 (there is no a_0). According to Lemma 4.2, any $2N + 1$ adjacent nodal intervals in $[-1, 1]$ that are at least distance $\delta > 0$ from the endpoints ± 1, all have comparable lengths to each other, with a multiplicative factor $1 + O(N/n\delta) = 1 + O((\log n)/n\delta)$. In particular, if n is large enough (depending on δ), then the roots of T_n contained in G_j satisfy the renormalized estimate (3.4). Let η_j denote the smallest length of a nodal interval for T_n inside G_j. Note that $\eta_j \leq |G_j|/(2N + 1)$, and that we can make $(2N+1)\eta_j/|G_j|$ as close to 1 as we wish by taking n sufficiently large. In other words, the roots of T_n inside G_j are as evenly spread as we wish, if n is large enough.

• **Step 4:** If $1 \leq M \leq N$, we say a set of $2M$ points $\{b_j\}_{|j|=1}^M \subset G_j$ is admissible if

(1) $\{b_{-N}, \dots, b_{-1}\}$ and $\{b_1, \dots b_N\}$ are each contained in disjoint subintervals J_-, J_+ of G_j of length at most $\eta_j/10$, and
(2) $\frac{1}{2}(b_{-k} + b_k) = c_k := \frac{1}{2}(a_{-k} + a_k)$ for $k = 1, \dots, M$.

We start with $M = N$ and claim we can choose an admissible set $\{b_{-1} < \cdots < b_{-N} < b_N < \cdots < b_1\}$ so that $J_- \subset G^L_j$ (the leftmost nodal interval) and $J_+ \subset G^R_j$ (the rightmost). Note that the ordering of the b_j's is different than for the a_j's. To see that we can meet the two required conditions for admissibility if n is sufficiently large, observe that by Lemma 4.2, the $\{a_k\}$ are as evenly spaced as we need, and thus all the c_k's are as close to the center of G_j as we wish (e.g., within $\eta_j/100$ of the center). This implies that we can place all the points b_k within $\eta_j/20$ of the centers of G_j^L or G_j^R. We denote these centered intervals of length $\eta_j/10$ as J_- and J_+.

• **Step 5:** If we perturb T_n by moving only the points $\{a_k\} \subset G_j$ to the points $\{b_k\} \subset G_j$, for a single j, then by (3.6) the new polynomial q_j has a node that is at least $10e^N$ larger than the original one. By Lemma 4.3 the perturbed polynomial has integral over $I^n_{k_j}$ that is at least $2e^N|I^n_{k_j}| \geq 2\exp(4\log n)4/n^2 \geq 8n^2$. Since the perturbed polynomial q_j has the same sign over its entire central node, the area of this central node of q_j is at least the integral of q_j over the subinterval $I^n_{k_j}$, and hence it satisfies the same lower bound.

• **Step 6:** We can obtain smaller areas over the central nodal interval by translating each point b_k by the same amount towards the center of G_j. Clearly such a translation preserves the distance between the points, so they still form two clusters of diameters at most $\eta_j/10$. The first contact between the b_k's and a_k's occurs when b_N hits a_N, (and b_{-N} hits a_{-N} at the same time). After this point, we stop moving b_N and b_{-N}, but keep translating the remaining points towards the center of G_j. Note that the remaining points form a $(N-1)$-admissible set, since they are still as tightly clustered as before (more so, since a point has been removed from each cluster). When any b_k reaches the corresponding point a_k, we stop moving it (and b_{-k}), but continue to move the remaining clusters. We finish when b_{-1} and b_1 reach a_{-1} and a_1 respectively. At that point we have returned to the original Chebyshev polynomial, which is bounded above by 1 on G_j, and hence has integral over this interval of at most $|G_j|$. Since the perturbed polynomial changes continuously with these movements, we can attain a node with any area between $|G_j| \leq 4\pi^2(\log n)/n$ and $2e^N|I^n_{k_j}| \geq 8n^2$. The roots $\{b_j\}$ that have not been matched with the corresponding a_j still form a M-admissible set for some $1 \leq M \leq N$.

We choose root positions so that the large central node in G_j of the perturbed polynomial has area $\epsilon \cdot e^{N/2} \simeq \epsilon \cdot n^2$, where $\epsilon > 0$ was the jump size used to define g. We can do this as long as $|G_j| < \epsilon$ and $e^N|I^n_{k_j}| \geq \epsilon \cdot e^{N/2}$. By Lemma 4.1, the first condition holds if n is large enough. The second condition holds if $|I^n_{k_j}| \geq \epsilon \cdot e^{-N/2}$. Again by Lemma 4.1, the nodal intervals I^n_k of T_n have length $|I^n_k| \geq \frac{4}{n^2}$, so this is true if $\frac{4}{n^2} \geq \epsilon \cdot e^{-N/2}$, or equivalently (since $\epsilon \leq 1$), if

$$N \geq 2\log\frac{\epsilon n^2}{4} = 4\log n - 2\log 4 + 2\log\epsilon > 4\log n.$$

In particular, the node is large enough for our choice $N = \lceil 4\log n \rceil$.

• **Step 7:** Let q be the polynomial obtained by making this perturbation in every G_j, for $j = 1, \ldots, K$, and define $p(x) = \int_{-1}^x e^{-N/2} \cdot q(t)\,dt$. We claim that the distortion function R_j corresponding to moving the roots from $\{a_j\}$ to $\{b_j\}$ in G_j satisfies

$$|R_j(x)| \leq 1 \quad \text{for} \quad x \in [-1,1] \setminus J_j. \tag{4.1}$$

This is easy for x outside G_j, because by Part (2) of Corollary 3.3, we have $0 < R_j(x) < 1$, for x outside $[b_{-1}, b_1]$, and hence outside G_j.

On the two components of $G_j \setminus J_j$, the argument is only slightly more involved. We can use Part (3) of Corollary 3.3 because of the condition we imposed in Step 3 that the roots $b_1 < \cdots < b_M$ are an M-admissible set. Since $|b_k - a_k| \geq \eta_j$ for $1 \leq k \leq M$, we can deduce that the distortion function corresponding to moving the pair a_k, a_{-k} to b_k, b_{-k} satisfies $|R| < 1$ on an interval of length at least $\eta_j/4$ to the left of b_k and to the right of b_{-k}. These intervals, together with $\{x < b_{-k}\}$ and $\{x > b_k\}$ contain all the points $\{b_j\}$ and thus cover all of $G_j \setminus J_j$ (recall that $J_j = (b_{-1}, b_1)$). This proves (4.1). Therefore $|p'| \leq e^{-N/2}$ in the set $X = [-1, 1] \setminus \cup_{j=1}^{K} J_j$ and hence the total variation of p over all the components of X is less than $2e^{-N/2}$. In particular, p is very close to constant on each connected component of X.

• **Step 8:** On each G_j, q has a large central node where it equals q_j (the perturbation due to perturbations inside G_j only) multiplied by the distortion due to each of the 2-point perturbations in the other intervals G_k, $k \neq j$. The distortion on G_j due to the perturbations in G_k is at most $1 + d^{-2}$ where $d := \operatorname{dist}(G_j, G_k)/|G_k|$. If we keep ϵ (and hence K) fixed, then $\max_{1 \leq k \leq K} |G_k|$ tends to zero with n, but $\operatorname{dist}(G_j, G_k) \geq \delta/2$, independent of n. Thus d tends to infinity as n tends to infinity.

Increasing n if necessary, we can assume that the distortion on G_j due to perturbations in other G_k's is as close to 1 as we wish. To be a little more precise, We have $d \simeq \delta/n \log n$ so the distortion due to all the 2-point distortions in G_j in some different G_k is bounded above by 1 and below by

$$(1 - \frac{1}{d^2})^{O(N)} = (1 - O(\frac{\delta^2}{n^2 \log^2 n}))^{O(\log n)}.$$

By taking logarithms, it is easy to check the right-hand side tends to 1 as $n \nearrow \infty$.

Thus the integral of p' over G_j is as close to $\pm\epsilon$ as we wish, say within ϵ/K, if n is large enough. Then the anti-derivative $p = \int p' = \int e^{-N/2} q$ equals g with an error of at most $2e^{-N/2} + \epsilon$. The first term, $2e^{-N/2}$, is due to the intervals between the $\{G_j\}$, and the second term, ϵ, is due to adding up at most K errors of size ϵ/K due to the distortions of the large nodes. By taking n (and hence N) large enough, we see that we can take $\sup_{[-1,1]} |f - p| \leq 2\epsilon$.

• **Step 9:** It is now easy to check that we can choose a sequence $\{p_m\}_1^\infty$ that forms a singular sequence converging to the continuous function f. Each p_m will have a derivative that is a multiple $e^{-N_m/2}$ of a perturbation q_m of the Chebyshev polynomial T_{n_m}, with $N_m = \lceil 4 n_m \rceil$ and the degrees n_m growing as quickly as we wish. Such polynomials were constructed in Steps 1 to 8. Each p_m approximates a step function g_m with some number K_m of "steps", as described above, and we may suppose that $\{G_j^m\}_{j=1}^{K_m}$ are the intervals where we performed the $2N_m$-point perturbations on the Chebyshev polynomial T_{n_m}. Let $G^m = \cup_{j=1}^{K_m} G_j^m$ be the union of these intervals. Note that $|p_m'| \leq e^{-N/2}$ off G^m. The length of G^m is

$$|G^m| = \sum_j |G_j^m| \le K_m(2N_m+1)\max_k |I_k^{n_m}| = O\left(\frac{K_m \log n_m}{n_m}\right).$$

We are free to choose a sequence $\{p_m\}_1^\infty$ so that K_m and n_m independently grow as quickly as we wish, so we choose it so that $K_m \ge m$ and $n_m \ge K_m^3$. Then

$$|G^m| = O\left(\frac{K_m \log n_m}{n_m}\right) = O\left(\frac{K_m \log K_m}{K_m^3}\right) = O\left(\frac{\log K_m}{K_m^2}\right) = O\left(\frac{\log m}{m^2}\right),$$

since $(\log x)/x$ and $(\log x)/x^2$ are both decreasing for $x \ge e$. This is summable over $m \ge 1$, so by the Borel-Cantelli lemma, almost every point of $[-1,1]$ is in only finitely many of the sets $\{G^m\}_1^\infty$. Thus $|p'_m| \to 0$ almost everywhere. This proves the first part of Theorem 1.2: every real-valued, continuous function f can be uniformly approximated by singular sequence of polynomials with only real critical points.

• **Step 10:** To prove the second part of the theorem, we need to show that if f is increasing, then we can choose $p' \ge 0$ everywhere on $[-1,1]$. This is fairly simple: replace q in the proof above by a q^2 and choose the points $\{b_j\}$ to represent pairs of roots that move together. As before, we can choose the new root locations so that $\int_{J_j} q^2 = e^{N/2}$. We then finish the proof as before. □

The proof of Theorem 1.2 given above shows that any sum of finite, real-valued point masses on $[-1,1]$ can be weakly approximated by a polynomial with only real roots. If the point masses are all positive, then we can take the polynomial to be nonnegative. Finite sums of point masses are weakly dense in all finite measures on $[-1,1]$, so we obtain the following consequence.

Corollary 4.4. *If μ is a finite Borel measure on $[-1,1]$, then there is a sequence of real polynomials $\{p_n\}$ with only real zeros so that p_n (restricted to $[-1,1]$) converges to μ weakly. If μ is positive, the polynomials $\{p_n\}$ can be chosen to be non-negative.*

5. Approximation by polynomials with large derivatives

In the previous section, we constructed polynomials that have large positive or negative spikes near specified locations, but that are small elsewhere, so that their antiderivatives approximate a step function. Thus we could uniformly approximate any continuous function f by a sequence of polynomials whose derivatives tend to zero pointwise almost everywhere. In this section, we want to construct approximating polynomials whose derivatives tend to $+\infty$ almost everywhere. Instead of approximating step functions, the graphs of our polynomial approximants will resemble "sawtooth" functions, i.e., functions that are piecewise linear, and have large positive slope on intervals that partition $[-1,1]$, but that have large downward jumps at the endpoints of these intervals. See Figs. 11 and 12 in Section 6 for such approximations of $f(x) = |x|$ and $f(x) = \cos(2\pi x)$.

We will need the following estimate that roughly holds because the nodes of a Chebyshev polynomial closely resemble rescalings of the nodes of $\cos t$:

$$\frac{1}{2} \le \frac{1}{|I_k^n|} \int_{I_k^n} |T_n|^2(t)\, dt \le \frac{1}{2} + \frac{\pi^2}{24n^2}, \tag{5.1}$$

where I_k^n is a nodal interval of T_n. Equation (5.1) is Equation (3.3) in [1] (there is a typo in that equation omitting the $1/|I_k^n|$ in front of the integral, but the proof of the equation with this additional normalization is correct). From (5.1) it is easy to deduce that $p(x) = \int_{-1}^{x} T_n^2(t)\, dt$ uniformly approximates the linear function $(x+1)/2$ on $[-1,1]$ as n tends to infinity.

All the nodes of T_n^2 are positive, but if we separate a double root of T_n^2 into two separate single roots, we introduce a single negative node between these roots. By moving these two simple roots further and further apart, we can create a very large negative node (we also move some of the double roots of T_n^2 to make room). The anti-derivative p of this perturbed polynomial q will look linear with slope $1/2$ sufficiently far from the perturbed roots, but it will have a sudden drop between the two simple roots; the size of the drop depends on the area of the negative node. By replacing T_n^2 by a large positive scalar multiple of itself, and by placing throughout $[-1,1]$ very large negative nodes, we will be able to uniformly approximate any continuous function f by a polynomial with the "sawtooth" structure described above. See Fig. 7 for a perturbation creating several negative nodes, all of the same size. In this figure, the negative nodes are too small to counteract the effect of the smaller, but more numerous positive nodes, and the anti-derivative resembles a linear function with positive slope. In Fig. 8, we have more carefully selected the negative nodes to balance the positive ones, and the resulting polynomial resembles a constant function, although its slope is very large at most points of $[-1,1]$. By choosing the size of the negative nodes more carefully, we can make the graph of the approximating polynomial approximate any Lipschitz function, as illustrated in Figs. 11 and 12.

Proof of Theorem 1.4. As in the Section 4, we will break the proof into a series of steps, although here we will omit listing them first, and simply start the proof. Several of the steps here are very similar to those used in Section 4, and we will refer back to those arguments when appropriate. We start with an important fact about Chebyshev polynomials, that will be used in Step 9 below.

Lemma 5.1. *Suppose n is a large, positive integer and consider T_n, the nth Chebyshev polynomial. Then $|\{x \in [-1,1] : |T_n(x)| \le \delta\}| \le \pi\delta$.*

Proof. Recall from Section 1 that $T_n = J((J^{-1})^n)$ where J is the vertical projection from the unit circle onto $[-1,1]$. In this formula, J^{-1} is interpreted as a 2 valued function on $[-1,1]$ taking x to $x \pm i\sqrt{1-x^2}$. Thus J^{-1} maps the interval $[-\delta,\delta]$ to two symmetric

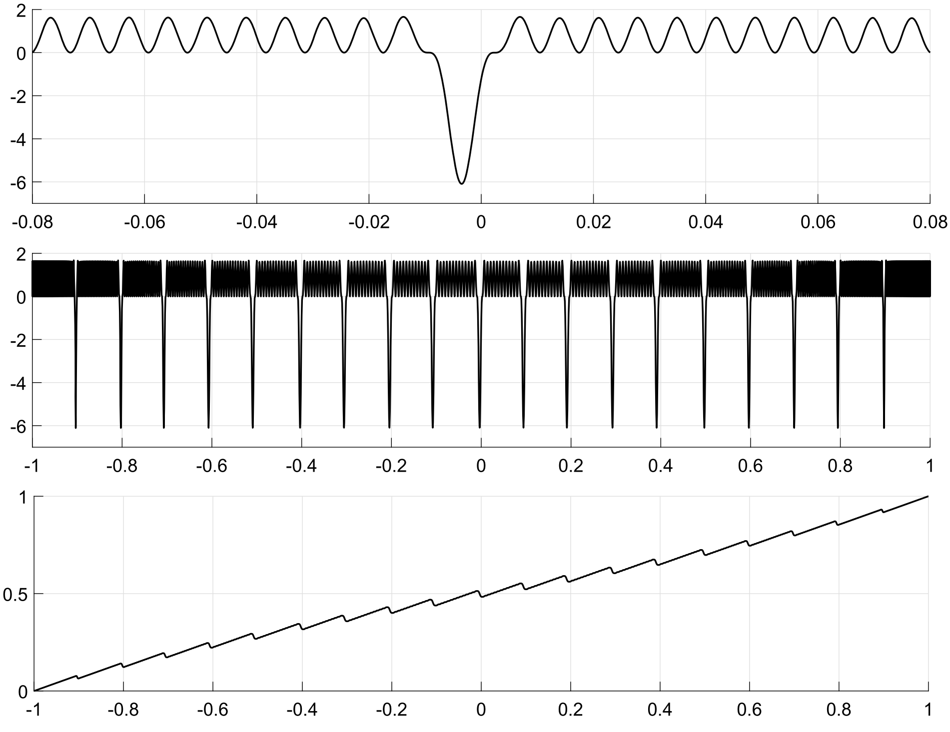

Fig. 7. Here we move six roots of q (three pairs of double roots) to form two roots of degree three each. Each such perturbation creates one large negative node. A single such node is shown, on top. Several, equal sized, nodes are shown in the middle picture. However, these negative nodes are "too small": the bottom figure shows the anti-derivative and the upward trend means the positive nodes dominate the negative ones. This is adjusted in Fig. 8.

arcs on $\mathbb{T}$, each of length $2\arcsin(\delta)$, and centered at $\pm i$. Then taking nth roots maps these two intervals to $2n$ intervals with the same total length, and the projection J maps these arcs 2-to-1 to n intervals, while deceasing the length of each interval. Thus the preimage of $[-\delta, \delta]$ under T_n has total length at most $2\arcsin(\delta) \leq \pi\delta$. $\square$

• **Step 1 (reduction to flat functions):** Since Lipschitz functions on $[-1, 1]$ are dense in continuous functions, it suffices to assume f is Lipschitz. Moreover, if Theorem 1.4 holds for a function, then it also holds for any positive scalar multiple of f, so we may further assume that f is 1-Lipschitz. Finally, if M is a large positive number, and if we can approximate f/M to within ϵ/M by a polynomial p so that $p' > M^{-1/2}$ except on a set of length $M^{-1/4}$, then $M \cdot p$ will approximate f to within ϵ and we have $(Mp)' > M^{1/2}$ except of a set of length $M^{-1/4}$. So it actually suffices to assume f is $(1/M)$-Lipschitz, and to approximate f by polynomials with these estimates. Fix a large value of M, say $M > 10$.

• **Step 2 (subdivide $[-1, 1]$):** Fix $\epsilon \in (0, 1/2)$ and set $J = 2\lfloor M/\epsilon \rfloor$. Note that $M/\epsilon \leq J \leq 2M/\epsilon$. Define $J + 2$ equally spaced points $-1 = s_0 < s_1 < \dots s_J < s_{J+1} = 1$. The distance between adjacent points is $2/(J + 1) < 2/J < 2\epsilon/M$. Since f is $(1/M)$-

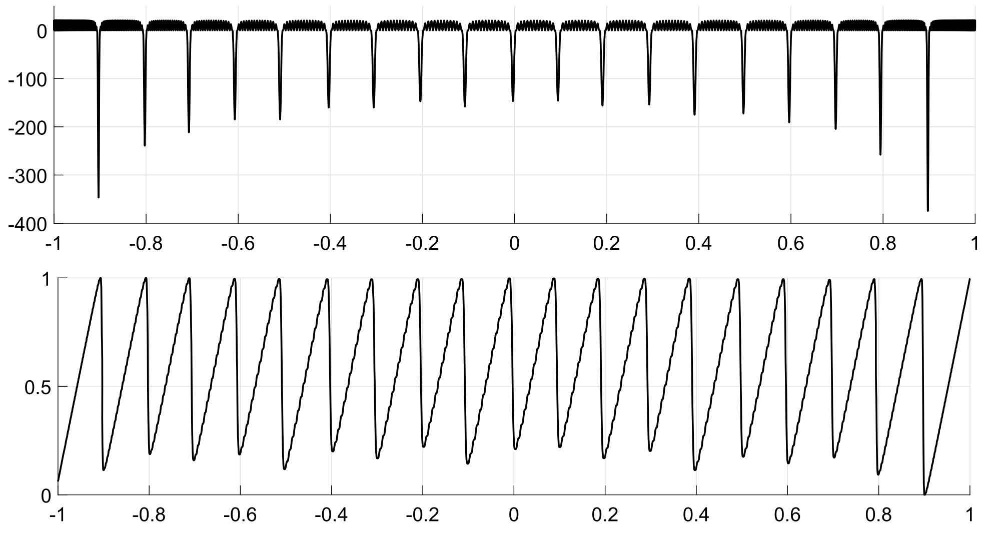

Fig. 8. Here we have adjusted the odd degree roots in Fig. 7 so each negative node has area equal to the mass of the following interval of positive nodes. The anti-derivative is shown at bottom. Nodes are larger near ± 1, to account for the shorter nodal intervals near the endpoints. The negative nodes take up about 20% of the length here, but this percentage can be made arbitrarily small by taking n and N larger.

Lipschitz, it varies by at most $(1/M)(2\epsilon/M) = 2\epsilon/M^2$ over each segment $S_j = [s_j, s_{j+1}]$ for $j = 0, \dots, J$. We will make perturbations of T_n^2 in very small intervals (size at most $O((\log n)/n)$) around the J points $s_1, \dots s_J$.

• **Step 3 (selecting the roots of the perturbed polynomial):** Suppose n is a large positive integer (chosen later depending on ϵ and M), and for each $j = 1, \dots J$, let r_j be a root of T_n that is closest to s_j; r_j is unique unless s_j happens to be the center of a nodal interval; in that case, let s_j be either endpoint of that interval. (As before, we suppress the dependence on n in the notation.) Set $N := 4\lceil \log n \rceil$, and let H_j be the union of the $N+1$ nodal intervals of T_n to either side of r_j. Thus H_j is the union of $2N+2$ nodal intervals, and there are $2N+1$ roots of T_n interior to H_j; each is double root of T_n^2. We let η_j be the minimal length of a nodal interval in H_j and we let H_j^L and H_j^R denote the leftmost and rightmost nodal intervals in H_j. As in Step 3 of the proof of Theorem 1.2, we may assume $(2N+2)\eta_j/|H_j|$ is as close to 1 as we wish, if n is large enough, depending only on J (hence only on ϵ and M).

We label the $2N+1$ roots of T_n inside H_j as $a_{-N} < \cdots < a_0 < \cdots < a_N$. Note that there is a point a_0 now, and that $a_0 = r_j$. (We ought to also label the a's with a superscript j, to indicate which H_j we are talking about, but this should always be clear from context, so we omit it to simplify notation.) As in Step 4 of Section 4, we define corresponding points $\{b_k\} \subset H_j$ that we move the roots at $\{a_k\}$ to (again, we omit adding a superscript j to the b's). However, there are $2N+2$ such points, instead of $2N+1$, since the double root of T_n^2 at a_0 will be split into two separate simple roots b_0^-, b_0^+. We assume these points satisfy

$$b_{-N} < \cdots < b_{-1} < b_0^- < b_0^+ < b_1 < \cdots < b_N.$$

As in Step 4 of the proof of Theorem 1.2, we also assume the points $\{b_k\}_{-N}^N$ are admissible in the sense that they satisfy two conditions:

(1) The set $\{b_{-N}, \dots, b_{-1}, b_0^-\}$ is contained in an interval J_- of length at most $\eta_j/10$ to the left of a_0, and the set $\{b_0^+, b_1, \dots, b_N\}$ is contained in an interval J_+ of the same length, but to the right of a_0.
(2) We have $\frac{1}{2}(b_{-k} + b_k) = c_k := \frac{1}{2}(a_{-k} + a_k)$ for $k = 1, \dots N$, and $\frac{1}{2}(b_0^- + b_0^+) = a_0$.

These conditions allow us to apply our estimates for 2-point distortion functions.

To start with, we choose admissible points $\{b_k\}$ so that $J_- \subset H_j^L$ and $J_+ \subset H_j^R$. We can do this for exactly the same reason as described in Step 4 of the proof of Theorem 1.2: if n is large enough, then the a_k's are as evenly distributed as we wish, so the c_k's are all as close to a_0 as desired, and hence the b_k's can be placed as close to the centers of H_j^L and H_j^R as needed.

• **Step 4 (defining perturbed polynomials):** For each k with $1 \le |k| \le N$ we will move the double root of T_n^2 at a_k to a double root at b_k. But for $k = 0$, the double root of T_n^2 at a_0 is split into two single roots, b_0^- and b_0^+. Note that b_0^- and b_0^+ are the closest to a_0 among the $\{b_k\}$. Since these are the only roots with odd multiplicity, these are the only points where the perturbed polynomial changes sign. Thus $J_j = (b_0^-, b_0^+) \subset H_j$ is a nodal interval of the perturbed polynomial, and it is the only nodal interval in H_j where this polynomial is negative. Let q_j denote the perturbation of T_n^2 obtained by moving the roots of T_n^2 to the points $\{b_k\}$ only in a single interval H_j. Let q be the polynomial obtained by perturbing the roots in all the intervals H_j, $j = 1, \dots J$. Then q has J negative nodes, one in each interval H_j for $j = 1, \dots, J$. Each perturbed polynomial q_j is equal to T_n^2 multiplied by $(4N+2)$ 2-point distortion functions. The polynomial q will be the product of T_n^2 and $J \cdot (4N+2)$ different 2-point distortion functions. Since each 2-point distortion function tends towards 1 as we move away from the perturbed roots, we will have $q \approx T_n^2$ away from $H = \cup_1^J H_j$. We will make this idea more precise below.

• **Step 5 (definition of p):** We define p as the anti-derivative of q so that $p(-1) = f(-1)$. Unlike the previous section, we do not multiply q a scalar (we used $e^{-N/2}$ in Section 4). Thus p will approximate a function of slope $1/2$ except near the intervals $\{H_j\}$, where it decreases rapidly due to the large negative nodes of q.

• **Step 6 (distortion far from the H_j's):** Set $\mu := \epsilon/M$. The distortion function for the perturbations inside H_j tends to 1 away from the interval H_j. In particular, for any $\mu > 0$, we can choose a $C < \infty$ (depending on μ) so that the distortion function satisfies $1 - \mu < R(x) < 1$ at points x that are outside $\widetilde{H}_j = (2C+1)H_j$ (where $(2C+1)H_j$ denotes the interval concentric with H_j but $(2C+1)$ times longer). We want to estimate the size of C in terms of μ.

By Part 4 of Lemma 3.3, the distortion function R due to a moving a single pair of points inside H_j satisfies $1 - C^2 \le R \le 1$ at points of $\widetilde{X} = [-1,1] \setminus \widetilde{H}_j$. Because

$\sum n^{-2}$ is summable, and because the intervals H_j are approximately evenly spaced, the product of distortion functions for moving one pair of roots in each H_j satisfies $1 - aC^{-2} \le R(x) \le 1$, for some fixed $a > 0$ and all $x \in \widetilde{X}$. Since $4N+4$ pairs of points are moved in each H_j, the distortion function corresponding to moving all $4N+4$ the roots satisfies $1-(aC^{-2})^{4N+4} \le R \le 1$. If we want $R > 1-\mu$, solving for C leads to the inequality $C^2 \gtrsim N/\mu = 4\lceil \log n \rceil \cdot M/\epsilon$. Therefore, if we fix C to be a sufficiently large multiple of $\lceil \sqrt{(\log n)M/\epsilon} \rceil$, then for $t \in \widetilde{X}$, we have

$$(1 - \frac{\epsilon}{M})T_n^2(t) \le q(t) \le T_n^2(t).$$

For two points $x < y$ in the same connected component of $\widetilde{X}$, we therefore have

$$(1 - \frac{\epsilon}{M}) \int_x^y T_n^2(t)\,dt \le p(y) - p(x) = \int_x^y q(t)\,dt \le \int_x^y T_n^2(t)\,dt,$$

and hence, by (5.1), as $n \nearrow \infty$ we have

$$|[p(y) - p(x)] - [\frac{1}{2}(y-x)]| \le \frac{\epsilon}{M} + O(n^{-2}).$$

• **Step 7 (distortion near H_j):** Outside the negative nodal interval J_j, the distortion function R_j due to the perturbations inside H_j satisfies $|R_j| \le 1$, for exactly the same reasons as in Step 7 of Section 4. Thus $|q| \le T_n^2$ outside these intervals, and so the variation of $p = \int q$ over $(2C+1)H_j \setminus J_j$ is bounded by the length of this set (since $|T_n| \le 1$), which is at most $(2C+1)|H_j| = O(\sqrt{(\log n)M/\epsilon}(\log n)/n)$. This clearly tends to zero as n tends to infinity (and ϵ and M are held fixed).

• **Step 8 (choosing the size of the negative nodes):** By Equation (3.6), that gives exponential growth of the nodes, we can choose the perturbation inside each H_j so that $\int_{J_j} q_j$ has a large negative value, up to size $-e^N|I^n_{k_j}|$, and we can achieve smaller values by translating the points $\{b_k\}$ towards the center of H_j, just as described in Step 6 in Section 4. In this case, we stop moving the double root at b_k, $k \ne 0$ when it reaches a_k. Thus $b_0^\pm$ are the last points to stop moving when they reach a_0; when this happens, we have returned to the unperturbed T_n^2. We choose the perturbation so that

$$\int_{J_j} q_j = -\frac{1}{2} \cdot |s_k - s_{k-1}| + (f(s_k) - f(s_{k-1})).$$

Note that this implies $\int_{J_j} q_j < -\frac{1}{4} \cdot |s_k - s_{k-1}|$ since f is $(1/M)$-Lipschitz and hence

$$f(s_k) - f(s_{k-1}) \le \frac{1}{M}(s_k - s_{k-1}) < \frac{1}{4}(s_k - s_{k-1})$$

if $M \geq 4$. A negative node of this size can be obtained if N is large enough, i.e., by (3.6) we need $e^N |I^n_{k_j}| \geq s_k - s_{k-1} = 2/(J+1)$. Similar to Step 6 of Section 4, this holds for our choice $N = 4\lceil \log n \rceil$, if n is so large that $\log n \geq \log M - \log \epsilon$.

• **Step 9 (conclusion):** We have now proven that for any $1 \leq j \leq J$, $p(s_j)$ approximates $f(s_j)$ as closely as wish, if n is large enough. Moreover, as noted in Step 3, f varies by at most ϵ/M^2 over $S_j = [s_j, s_{j+1}]$. On S_j, the polynomial p is decreasing (due to the node J_j), then increases by at most $1/J$ (since $p' \leq 1$ between J_j and J_{j+1}) and then decreases again (due to J_{j+1}). Therefore the variation of p over S_j is at most $O(1/J) = O(\epsilon/M)$ if n is large enough. Thus $|f - p| \leq O(\epsilon/M)$ on all of $[-1, 1]$ if n is sufficiently large.

We have also shown that $p' = q$ is larger than $M^{-1/2}$ except inside $\widetilde{H} = \cup_j \widetilde{H}$ and near the zeros of q that are outside this set. Outside $\widetilde{H}$, we have $|q| \geq (1-\mu)|T_n| \geq |T_n|/2$, so by Lemma 5.1, $|q| \geq M^{-1/2}$ except on a union of intervals of total length at most $O(M^{-1/4})$. On the other hand, $\cup_j \widetilde{H}_j$ has total length bounded by

$$O\left(J \cdot (2C+1) \cdot N/n\right) = O\left((M/\epsilon) \cdot \sqrt{(\log n)M/\epsilon} \cdot (\log n)/n\right),$$

(this is J, times the total number of nodal intervals in $\cup_j \widetilde{H}_j$, times the maximal possible length of a nodal interval, as given by Lemma 4.1). Clearly this estimate tends to zero as n increases, so it is less than $M^{-1/4}$ for n large enough. To complete the proof of Theorem 1.4, take a sequence of polynomial approximants $\{p_n\}$ constructed as above, and so that the corresponding values $\{M_n\}$ satisfy $\sum M_n^{-1/4} < \infty$. Then applying the Borel-Cantelli lemma, as we did in the proof of Theorem 1.2, gives the result. $\square$

Our methods can be adapted to prove the following result.

Theorem 5.2. *Suppose E_1, E_2, and E_3 are disjoint measurable sets in I. If f is real-valued and continuous on I, then there is a sequence of polynomials $\{p_n\}$ with only real critical points so that*

(1) *on E_1, $\{p'_n\}$ converges almost everywhere to 0,*
(2) *on E_2, $\{p'_n\}$ converges almost everywhere to $+\infty$*
(3) *on E_3, $\{p'_n\}$ converges almost everywhere to $-\infty$.*

We shall leave the details to the reader, but the basic idea is as follows. Fix $\epsilon > 0$ and split $[-1, 1]$ into finitely many intervals so that f varies by less than ϵ over each. Within each interval that intersects E_2 in more than three quarters of its length, we perturb the roots to form large positive nodes; in the intervals that hits E_3 at least three quarters of their length, we form large negative nodes. In both cases the nodes are chosen with the same very large area (positive or negative). We then rescale the perturbed function so the absolute mass of each interval is approximately its length. The remaining intervals each hit E_1 in at least half their length. On these, the renormalized polynomial is very close to zero. Finally, we introduce large nodes near the endpoint of intervals whose

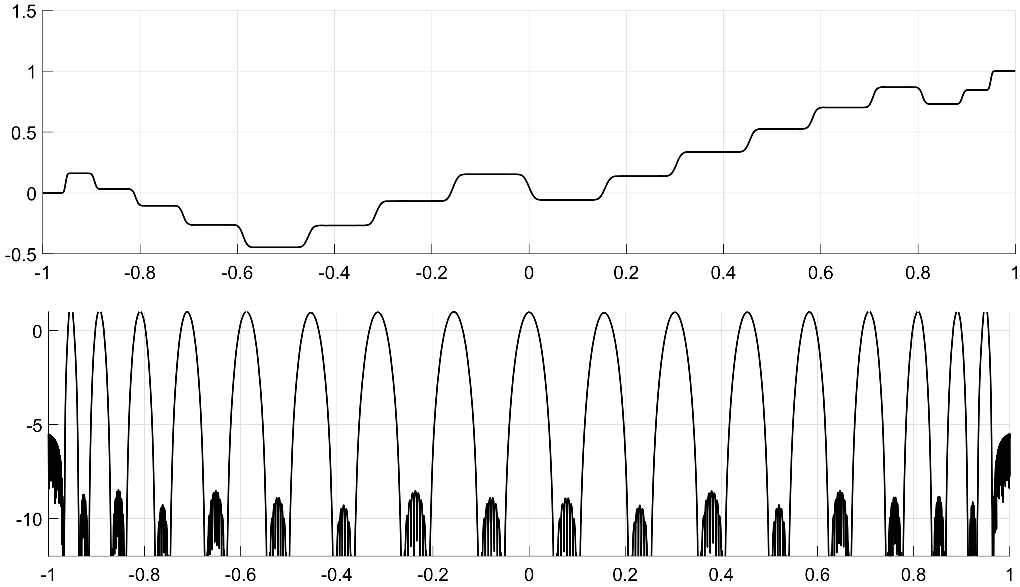

Fig. 9. A degree 500 polynomial p approximating a step function and a plot of $\log_{10}|p'|$. The polynomial has large slope near the "jumps" but small slope elsewhere. We have moved 5 pairs of points per jump.

total mass is the sum of the change in the target function f over the interval to the left, minus the mass (positive or negative) of the interval to the left. The anti-derivative p of resulting polynomial will approximate f uniformly and p' will be close to 0, $-\infty$ or $+\infty$ on large measure on the three specified sets respectively. Taking appropriate sequences formed in this way and applying the Borel-Cantelli theorem proves Theorem 5.2.

6. Some numerical examples

The proof of Theorem 1.2 is illustrated in Fig. 9. The top picture shows a degree 500 polynomial p approximating a step function. The bottom picture shows a graph of $\log_{10}|p'|$ with the vertical range limited to $[-12, 1]$. The polynomial is approximately 10-Lipschitz, but outside the intervals where we move the roots of the Chebyshev polynomial, $|p'|$ is everywhere less than 10^{-3}, and is less than 10^{-4} except near ± 1. At each "jump" of the function we have moved 10 roots.

To simplify the computation of this example, we made the new nodes about the same height, without adjusting for the width of the nodal interval. Hence $|p'|$ has about the same maximum at each jump, but the steps of p are smaller near the endpoints because the nodal intervals are shorter. As described in the proof of Theorem 1.2 above, this can be adjusted so that the jumps all have same height.

Fig. 10 shows an approximation to a Cantor singular function. We have squared T_{450} and moved some roots to new roots of degree six. The steps have derivatives that are almost four orders of magnitudes larger than in the intervening intervals.

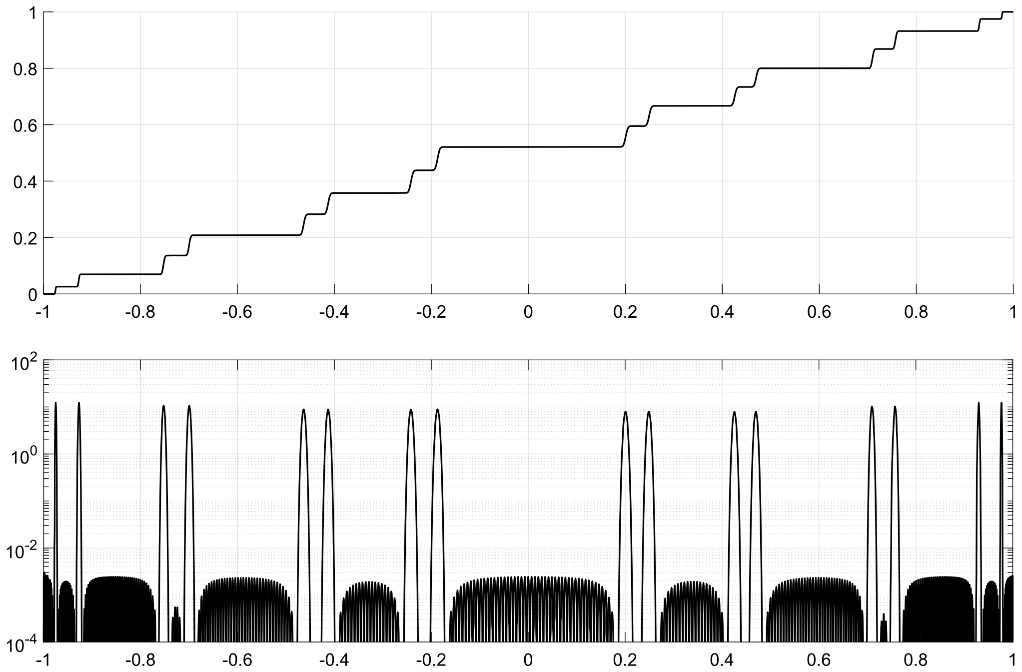

Fig. 10. An increasing, degree 900 polynomial approximating a Cantor singular function. Below is a plot of the derivative with a logarithmic vertical scale.

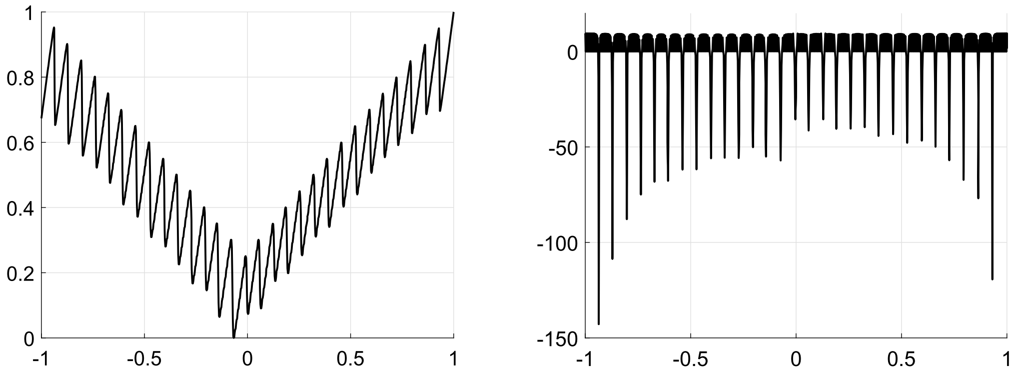

Fig. 11. A polynomial p of degree 1600 approximating $f(x) = |x|$, and a graph of p'. The derivative p' is given by squaring T_{800} and then moving certain groups of adjacent roots to form large negative nodes that mostly cancels the positive nodes.

To prove Theorem 1.4, we constructed polynomials $q = p'$ that are bigger than some large constant M on most of $[-1, 1]$ but have even larger negative nodes supported on very small length. Thus the anti-derivative p looks like a "sawtooth", i.e., p resembles a function of the form $Mx - g(x)$ where g is a step function. See Figs. 11 and 12 for some examples where we have implemented the idea to approximate $|x|$ and $\cos 2\pi x$ using two polynomials of degree 1600. The approximations are very rough; even higher degrees are

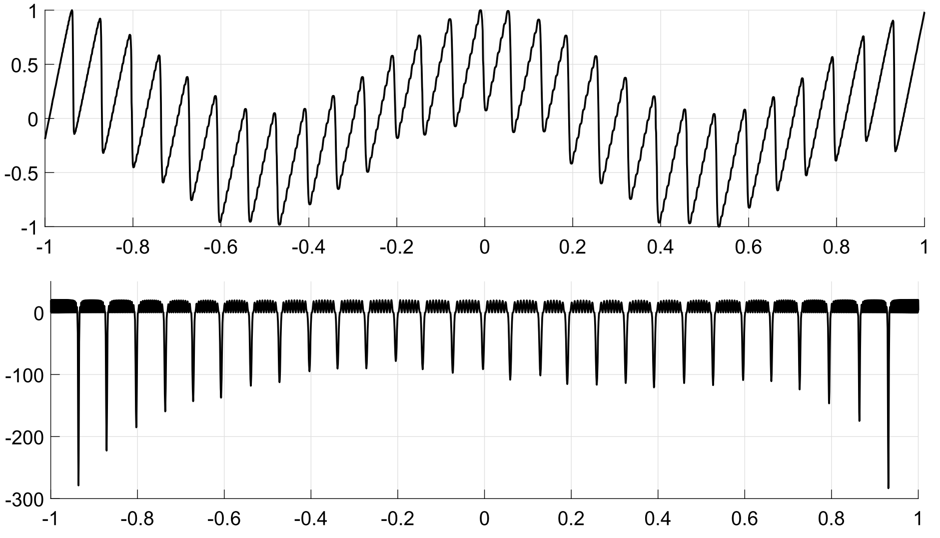

Fig. 12. A polynomial approximation of $\cos(2\pi x)$, and a graph of p'. The derivative is large and positive on a set of large measure (close to measure 1) but is balanced by even larger negative spikes supported on small length.

necessary to get close approximations by polynomials with large, positive derivatives on a large set.

References

[1] D.L. Bishop, Approximation by polynomials with only real critical points, Rev. Mat. Iberoam. 40 (6) (2024) 2251–2290.

[2] C.J. Bishop, K. Lazebnik, A geometric approach to polynomial and rational approximation, Int. Math. Res. Not. 12 (2024) 9936–9961.

[3] J.E. Brown, V.F. Powell, A result on real polynomials with real critical points, J. Anal. Appl. 5 (1) (2007) 41–52.

[4] J.G. Clunie, A.B.J. Kuijlaars, Approximation by polynomials with restricted zeros, J. Approx. Theory 79 (1) (1994) 109–124.

[5] J.G. Clunie, Convergence of polynomials with restricted zeros, in: Functional Analysis and Operator Theory, New Delhi, 1990, in: Lecture Notes in Math., vol. 1511, Springer, Berlin, 1992, pp. 100–105.

[6] T.A. Driscoll, N. Hale, L.N. Trefethen, Chebfun Guide, Pafnuty Publications, 2014.

[7] A.L. Epstein, Symmetric rigidity for real polynomials with real critical points, in: Complex Manifolds and Hyperbolic Geometry, Guanajuato, 2001, in: Contemp. Math., vol. 311, Amer. Math. Soc., Providence, RI, 2002, pp. 107–114.

[8] A. Hinkkanen, I. Kayumov, On critical values of polynomials with real critical points, Constr. Approx. 32 (2) (2010) 385–392.

[9] J. Korevaar, C. Loewner, Approximation on an arc by polynomials with restricted zeros, Proc. K. Ned. Akad. Wet. 67, Indag. Math. 26 (1964) 121–128.

[10] O. Kozlovski, W. Shen, S. van Strien, Rigidity for real polynomials, Ann. Math. (2) 165 (3) (2007) 749–841.

[11] E. Laguerre, Sur les fonctions du genre zero et du genre un, C. R. Séances Acad. Sci. (1882) 174–177.

[12] D.E. Marshall, Complex Analysis. Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 2019.

[13] G. Pólya, Über annäherung durch polynome mit lauter reelen wurzeln, Rend. Circ. Mat. Palermo 36 (1913) 279–295.

[14] K. Weierstrass, Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen, Sitz.ber. Königlich Preuss. Akad. Wiss. Berl. (1885).