Yale University
Department of Computer Science

Polynomial Interpolation, Threshold Circuits, and the Polynomial Hierarchy

Richard Beigel

YALEU/DCS/TR-843
January 22, 1991

Polynomial Interpolation, Threshold Circuits, and the Polynomial Hierarchy

Richard Beigel*
Yale University

Abstract

Toda [13] has shown that the polynomial hierarchy is contained in P^{PP}. It is natural to ask whether the polynomial hierarchy is in fact contained in PP. Along these lines, it has been shown [2] that $P^{NP[\log]}$ is contained in PP. However, a lower bound of Minsky and Papert [8] implies that Σ_2^p is not contained in PP relative to an oracle [5]. Thus we ask how much of the polynomial hierarchy is contained in PP.

We construct an oracle relative to which $P^{NP[f(n)]}$ is contained in PP if and only if $f(n) = O(\log n)$, so the results of [2] are optimal in a relativized world. In particular, relative to this oracle, Δ_2^p is not contained in PP. Our oracle is also the first relative to which P_{tt}^{NP} is properly contained of P_T^{NP}, or in the terminology of Wagner's refined polynomial hierarchy [15], θ_2^p is properly contained in Δ_2^p.

Our construction depends on a new lower bound for perceptrons, which is interesting in its own right. We construct a predicate that is computable by a small perceptron, but which requires exponentially large weights. This lower bound depends in turn on a fundamental property of polynomials: if p is bounded on the domain $\{1, \ldots, m\}$ then the coefficients of p must be small as a function of m.

1. Introduction

Recently Toda [13] proved that the polynomial hierarchy is contained in P^{PP}, and Beigel, Hemachandra, and Wechsung [2] proved that $P^{NP[\log]}$ is contained in PP. Many people have asked whether those two results could be extended to show that the polynomial hierarchy is contained in PP. As a corollary to Minsky and Papert's [8] "one-in-a-box" theorem, where they construct a depth-2 AC^0 predicate that is not recognized by any perceptron with order less than $\frac{1}{2}\sqrt{n}$, Bin has shown that there is

*Dept. of Computer Science, 51 Prospect Street, P.O. Box 2158, Yale Station, New Haven, CT 06520-2158. Email: beigel-richard@cs.yale.edu. Research performed while a visiting scholar at Stanford University. Supported in part by NSF grants CCR-8808949 and CCR-8958528.

an oracle A such that $(\Sigma_2^p)^A \not\subseteq \mathrm{PP}^A$. Thus it is unlikely that current techniques will be extended to show that $\mathrm{PH} \subseteq \mathrm{PP}$.

The result of [2] has been improved in [7] where it was shown that $\mathrm{P}^{\mathrm{C_=P[log]}} \subseteq \mathrm{PP}$ and in [4] where it was shown that $\mathrm{P}^{\mathrm{PP[log]}} \subseteq \mathrm{PP}$. Toda's result [12] has been improved independently in [14] and [11] where it was shown that $\mathrm{PH} \subseteq BP \cdot \mathrm{PP}$.

It is natural to ask whether $(\Delta_2^p)^A$ is contained in PP. We construct an oracle A such that $(\Delta_2^p)^A \not\subseteq \mathrm{PP}^A$. In fact, we obtain $\mathrm{P}^{\mathrm{NP}^A[f(n)]} \not\subseteq \mathrm{PP}^A$ unless $f(n) = O(\log n)$, showing that the result of [2] is tight in relativized worlds.

The structure of the proof is as follows: We define a test language

$$\text{ODD-MAX-ELEMENT}^A = \{0^n : \max(A \cap \{0,1\}^n) \text{ ends in a } 1\},$$

which belongs to $\mathrm{P}^{\mathrm{NP}^A}$, and we define a related language ODD-MAX-BIT, which is the set of strings whose rightmost 1 is in an odd-numbered position.

We assume some familiarity with circuits. A weighted *threshold*-gate with weights $w_1, \ldots, w_n$ outputs 1 on inputs $x_1, \ldots, x_n$ iff $\sum w_i x_i > 0$. All weights must be integers. A *perceptron* is a circuit with a weighted threshold gate at the root and AND-gates at the remaining level. The *order* of a perceptron is the maximum fanin of its AND-gates. The *weight* of a perceptron is the maximum absolute value of the weights on the inputs to its threshold gate. The *size* of a perceptron is the number of AND-gates it contains. Perceptrons are an important computational model, which is used in practice, and which has been studied in [8, 4, 1, 11, 3].

The correspondence between oracle Turing machines and circuits is as in [6]. We will just sketch the basic idea. We construct an oracle A such that $\text{ODD-MAX-ELEMENT}^A \notin \mathrm{PP}^A$ by an initial segment argument. In order to defeat a polynomial-time probabilistic oracle TM M we choose m such that $A \cap \{0,1\}^m$ is as yet completely undefined. By convention we assume that a computation of M includes the oracle answers. Fix an input 0^m. Let $n = 2^m$. We will construct a perceptron C of size $2^{\mathrm{polylog}\, n}$, weight 1, and order $\mathrm{polylog}\, n$ that simulates M. Its input consists of $n = 2^m$ bits: the characteristic sequence of $A \cap \{0,1\}^m$. For each of the $2^{\mathrm{polylog}\, n}$ computations of M, we construct an AND-gate that verifies the oracle answers in the computation; each such AND-gate has fanin $m^{O(1)} = \mathrm{polylog}\, n$. If a computation accepts, then we give its AND-gate weight $+1$; if it rejects, then we give its AND-gate weight -1. The perceptron C accepts the characteristic sequence of $A \cap \{0,1\}^m$ if and only if M accepts 0^m when using oracle A. We choose $A \cap \{0,1\}^m$ so that C accepts or rejects incorrectly. The construction fails only if there is a family of perceptrons having size $2^{\mathrm{polylog}\, n}$, weight 1, and order $\mathrm{polylog}\, n$ which compute the predicate ODD-MAX-BIT.

It is easy to construct a family of perceptrons having size $2^{\mathrm{polylog}\, n}$ and order $\mathrm{polylog}\, n$ which compute the predicate ODD-MAX-BIT. However we show that any such circuits require exponentially large weights. Thus the desired oracle must exist. The techniques are novel, because all previous lower bounds for perceptron size hold regardless of weights. (Although there are examples in [8, Sections 10.1-4] where large weights are proved necessary, these either involve perceptrons whose size is large anyway, or else very contrived computational models.)

Our circuit lower bound depends on a technical result concerning polynomial interpolation, which we prove in the appendix. Let p denote a real polynomial having degree d. It is well known that if p takes the value 0 at $d+1$ distinct points then all of its coefficients must be 0. We prove an analogous result for polynomials that are approximately 0 at many points: If p is bounded in absolute value by a small constant at the points $1, \ldots, m$ where m is quite large compared to d, then all of the coefficients of p, except the constant term, are extremely small. Hence, we may conclude that $p(1) - p(0)$ is very small.[1]

2. Threshold Circuits

Let $\max(S)$ denote the lexically maximum string belonging to the finite set S.

Definition 1.

- ODD-MAX-ELEMENT$^A = \{0^n : \max(A \cap \{0,1\}^n)$ ends in a $1\}$.
- ODD-MAX-BIT is the set of all strings over $\{0,1\}^*$ whose rightmost 1 is in an odd-numbered position, i.e., the set of strings of the form $x10^k$ where the length of x is even.

Let $\mathrm{P}^{\mathrm{NP}^A[f(n)]}$ denote the class of languages accepted by a deterministic polynomial-time bounded oracle Turing machine that is allowed at most $f(n)$ queries to an NP^A oracle. The following proposition is standard.

Proposition 2.

- *If, for every oracle A,* ODD-MAX-ELEMENT^A *belongs to* PP^A*, then n-bit instances of* ODD-MAX-BIT *can be decided by perceptrons having size* $2^{\mathrm{polylog}\, n}$*, weight* 1*, and order* $\mathrm{polylog}\, n$.
- *If, for every oracle A,* $\mathrm{P}^{\mathrm{NP}^A[f(n)]}$ *belongs to* PP^A*, then* $(2^{f(n)}-1)$*-bit instances of* ODD-MAX-BIT *can be decided by perceptrons having size* $2^{n^{O(1)}}$*, weight* 1*, and order* $n^{O(1)}$.

We say that a perceptron is in *clean* form if it contains no negations and no identical AND-gates. The following lemma is essentially due to Minsky and Papert [8].

Lemma 3. *If f is computed by a perceptron with size s, weight w, and order d, then f is computed by a perceptron in clean form with size $2^d s$, weight sw, and order d.*

[1] The conclusion that $p(1) - p(0)$ is very small may also be obtained by careful analysis of the proof of Lemma 2.2.2 in Mario Szegedy's doctoral dissertation [10].

Proof: For each AND-gate, replace each negated input $\overline{x}$ by $1 - x$, and replace "and" by multiplication. Expand using the distributive laws of arithmetic. Each term in the expansion is a conjunction of the inputs to C. The total number of terms obtained by expanding all AND-gates in this way is at most $2^d s$. Each term is contributed at most once per AND-gate, so when we collect terms, no weight is greater than sw in absolute value. ∎

Lemma 4. *If C is a perceptron in clean form having size s, weight w, and order d which recognizes* ODD-MAX-BIT $\cap \{0,1\}^n$ *then*

$$w \geq \frac{1}{s} 2^{(n-1)/48d^{10}}.$$

Proof: Let $x = x_1, \ldots, x_n$ denote the input to C. We identify the vector x with the set $X = \{i : x_i = 1\}$. By assumption, C accepts x iff $\max(X)$ is odd. Let T denote C's threshold gate. Each input to T depends on a set $S \subseteq \{1, \ldots, n\}$ such that $0 \leq |S| \leq d$. For each S let $w(S)$ denote the weight given to the corresponding input to T (0 if there is no such input).

Let

$$c(X) = \sum_{S \subseteq X, |S| \leq d} w(S)$$

denote the *total weight of* X. Then C accepts x iff $c(X) > 0$. Hence $c(\{1\}) \geq 1$. For $0 \leq i \leq d$, let

$$c_i(S) = \sum_{S \subseteq X, |S| = i} w(S)$$

denote the weight of X due to subsets of size i. Clearly

$$c(X) = \sum_{0 \leq i \leq d} c_i(X).$$

Now suppose that we have found X such that $c(X) = -W < 0$ and $\max(X) = i$. Let $m = 24d^{10}$, and let $M = \{i+1, i+3, \ldots, i+2m-1\}$. We will find $Y \subseteq M$ such that $c(X \cup Y) \geq 2W$. Similarly if $c(X) > 0$, we will find $Y \subseteq M$ such that $C(X \cup Y) \leq -2c(X)$. We start with $X = \{1\}$ and $W \geq 1$, and we iterate $(n-1)/48d^{10}$ times to obtain a set $X \subseteq \{1, \ldots, n\}$ with $c(X) \geq 2^{(n-1)/48d^{10}}$, so $w \geq \frac{1}{s} 2^{(n-1)/48d^{10}}$.

It remains to show that the desired set Y always exists. We will consider only the case $c(X) < 0$, because the other case is entirely similar. Fix X, i, m, and M as above. For each $Y \subseteq M$, we have $c(X \cup Y) \geq 0$. For each $S \subseteq M$ let

$$b(S) = \sum_{X \cup R = X \cup S, |R| \leq d} w(R)$$

denote the weight due to S. Let

$$u_k(Y) = \operatorname*{ave}_{S \subseteq Y, |S| = k} b(S)$$

denote the average weight due to a k-element subset of Y. Note for each $j \geq k$ that $u_k(M)$ is equal to $\text{ave}_{Y \subseteq M, |Y|=j}\, u_k(Y)$. Now

$$\begin{aligned} c(X \cup Y) &= \sum_{S \subseteq Y} b(S) \\ &= c(X) + \sum_{S \subseteq Y, 1 \leq |S| \leq d} b(S) \\ &= c(X) + \sum_{1 \leq k \leq d} \sum_{S \subseteq Y, |S|=k} b(S) \\ &= c(X) + \sum_{1 \leq k \leq d} \binom{|Y|}{k} u_k(Y). \end{aligned}$$

Therefore, for every $Y \subseteq M$ we have

$$\sum_{1 \leq k \leq d} \binom{|Y|}{k} u_k(Y) = c(X \cup Y) - c(X) \geq W. \tag{1}$$

If, for some $Y \subseteq M$ we have

$$\sum_{1 \leq k \leq d} \binom{|Y|}{k} u_k(Y) \geq 3W$$

then $c(X \cup Y) \geq 2W$ and we are done, so assume that for every $Y \subseteq M$ we have

$$\sum_{1 \leq k \leq d} \binom{|Y|}{k} u_k(Y) < 3W. \tag{2}$$

Recall that for $j \geq k$, $u_k(M)$ is equal to $\text{ave}_{Y \subseteq M, |Y|=j}\, u_k(Y)$. Therefore, by (1) and (2),

$$W \leq \sum_{1 \leq k \leq d} \binom{j}{k} u_k(M) < 3W.$$

Define a dth degree polynomial

$$p(z) = \sum_{1 \leq k \leq d} u_k(M) \binom{z}{k}.$$

Then we have $W \leq p(z) < 3W$ for $z = 1, 2, \ldots, m$. We can also expand $p(z)$ to obtain constants $a_1, \ldots, a_d$ such that

$$p(z) = \sum_{1 \leq k \leq d} a_k z^k.$$

By Lemma 11 we have

$$a_k \leq 2\left(\frac{2}{k}\right)^k 3Wd \left(\frac{d^9}{m + d/255 + 1}\right)^k$$

$$
\begin{aligned}
&\leq 12Wd\left(\frac{d^9}{m}\right)^k \\
&= 12Wd\left(\frac{d^9}{24d^{10}}\right)^k \qquad \text{since } m = 24d^{10} \\
&\leq 12Wd\left(\frac{1}{24d}\right)^k.
\end{aligned}
$$

Therefore

$$
p(1) = \sum_{1\leq k\leq d} a_k \leq \frac{1}{2}W\frac{1}{1-1/24d} \leq \frac{12}{23}W,
$$

which contradicts $p(1) \geq W$. ∎

Theorem 5. *If $f(n) \neq O(\log n)$ then there exists an oracle A such that $\mathrm{P}^{\mathrm{NP}^A[f(n)]}$ is not contained in PP^A.*

Proof: Suppose that $\mathrm{P}^{\mathrm{NP}^A[f(n)]} \subseteq \mathrm{PP}^A$. Then $(2^{f(n)}-1)$-bit instances of ODD-MAX-BIT can be decided by perceptrons with size $2^{n^{O(1)}}$, weight 1, and order $n^{O(1)}$. Therefore they can be decided by perceptrons in clean form with size $s = 2^{n^{O(1)}}$, weight $w = 2^{n^{O(1)}}$, and order $d = n^{O(1)}$. By Lemma 4, $w \geq \frac{1}{s}2^{(2^{f(n)}-1)/48d^{10}}$. Therefore,

$$
\begin{aligned}
\frac{1}{s}2^{(2^{f(n)}-2)/48d^{10}} &= 2^{n^{O(1)}}, \\
2^{(2^{f(n)}-2)/48d^{10}} &= s2^{n^{O(1)}}, \\
2^{(2^{f(n)}-2)/48d^{10}} &= 2^{n^{O(1)}}, \\
(2^{f(n)}-2)/48d^{10} &= n^{O(1)}, \\
2^{f(n)}-2 &= n^{O(1)}48d^{10}, \\
2^{f(n)}-2 &= n^{O(1)}, \\
2^{f(n)} &= n^{O(1)}, \\
f(n) &= O(\log n).
\end{aligned}
$$

∎

Corollary 6. *There exists an oracle A such that $\mathrm{P}^{\mathrm{NP}^A}$ is not contained in PP^A.*

Corollary 7 (Obtained independently by Bin [5]). *There exists an oracle A such that PH^A is not contained in PP^A.*

Corollary 8. *There exists an oracle A such that $(\theta_2^p)^A \subset (\Delta_2^p)^A$.*

Acknowledgments. I am grateful to Lane Hemachandra, Gerd Wechsung, Tomas Feder, Anna Karlin, Samuel Karlin, Bill Gasarch, John Gill, Joan Feigenbaum, and Jun Tarui for helpful discussions; to Allen Cohn, Anna Karlin, Jeff Westbrook, and Bob Floyd for their hospitality during my visit to Stanford; and to Martin Schultz, Michael Fischer, and Nick Reingold for providing the free time that made that visit possible.

References

[1] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. On the expressive power of voting polynomials. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, 1991. To appear.

[2] R. Beigel, L. A. Hemachandra, and G. Wechsung. On the power of probabilistic polynomial time: $P^{NP[log]} \subseteq PP$. In *Proceedings of the 4th Annual Conference on Structure in Complexity Theory*, pages 225–227, June 1989.

[3] R. Beigel, N. Reingold, and D. Spielman. The perceptron strikes back. YALEU/DCS/TR 813, Yale University, Dept. of Computer Science, 1990.

[4] R. Beigel, N. Reingold, and D. Spielman. PP is closed under intersection. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, 1991. To appear.

[5] F. Bin. Separating PH from PP by relativization. Preprint, 1990.

[6] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, April 1984.

[7] T. Gundermann, N. Nasser, and G. Wechsung. A survey of counting classes. In *Proceedings of the 5th Annual Conference on Structure in Complexity Theory*, pages 140–153. IEEE Computer Society Press, July 1990.

[8] M. L. Minksy and S. A. Papert. *Perceptrons.* MIT Press, Cambridge, MA, 1988. Expanded Edition. The first edition appeared in 1968.

[9] T. Rivlin, 1990. Personal communication.

[10] M. Szegedy. *Algebraic Methods in Lower Bounds for Computational Models with Limited Communication.* PhD thesis, The University of Chicago, Dec. 1990.

[11] J. Tarui. Randomized polynomials, threshold circuits, and the polynomial hierarchy. To appear, 1991.

[12] S. Toda. On polynomial-time truth-table reducibilities of intractable sets to p-selective sets. Manuscript, Apr. 1988.

[13] S. Toda. On the computational power of PP and $\oplus$P. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 514–519, 1989.

[14] S. Toda and M. Ogiwara. Counting classes are as hard as the polynomial-time hierarchy. Manuscript, 1990.

[15] K. W. Wagner. Bound query computations. In *Proceedings of the 3rd Annual Conference on Structure in Complexity Theory*, pages 260–277. IEEE Computer Society Press, June 1988.

Appendix: Polynomial Interpolation

We consider a polynomial p that is bounded on the domain $\{1,\ldots,m\}$ and we show that the coefficients of p must be very small. This is what one would expect, but the proofs are not trivial. Ted Rivlin [9] reports some improvements to our bounds.

Lemma 9. *Let* $p(x) = \sum_{0\le k\le d} a_k x^k$ *be a* d*th degree polynomial, and let* $c = \max_{1\le i\le m} |p(i)|$, *for some* $m \ge d+1$. *Then*

$$|a_k| \le c\left(\frac{4(d+1)}{k(m-d)}\right)^k 4^d.$$

Proof: We proceed by choosing $d+1$ equally spaced points between 1 and m and passing a dth degree polynomial through those points. Since that polynomial is unique, it must be equal to p. Then we estimate its coefficients.

Let μ be the unique multiple of $d+1$ in $[m-d, m]$. Let $\delta = \mu/(d+1)$. For $1 \le i \le d+1$ let $x_i = i\delta$, and let $c_i = p(x_i)$. By Lagrange's interpolation formula,

$$\begin{aligned}
p(x) &= \sum_{1\le i\le d+1} c_i \prod_{j\ne i} \frac{x-x_j}{x_i-x_j} \\
&= \sum_{1\le i\le d+1} c_i \prod_{j\ne i} \frac{x-j\delta}{(i-j)\delta} \\
&= \sum_{1\le i\le d+1} \frac{c_i}{(-1)^{d+1-i}(d+1-i)!(i-1)!\delta^d} \prod_{j\ne i}(x-j\delta) \\
&= \delta^{-d} \sum_{1\le i\le d+1} \frac{c_i}{(-1)^{d+1-i}(d+1-i)!(i-1)!} \prod_{j\ne i}(x-j\delta).
\end{aligned}$$

Therefore,

$$\begin{aligned}
|a_k| &\le c\delta^{-d} \sum_{1\le i\le d+1} \frac{1}{(i-1)!(d+1-i)!}\binom{d}{k} \prod_{k+1<j\le d+1} j\delta \\
&= c\delta^{-d} \sum_{1\le i\le d+1} \frac{1}{(i-1)!(d+1-i)!}\binom{d}{k} \delta^{d-k}\frac{(d+1)!}{(k+1)!} \\
&= c\delta^{-k} \sum_{1\le i\le d+1} \frac{1}{(i-1)!(d+1-i)!}\binom{d}{k} \frac{(d+1)!}{(k+1)!} \\
&= \frac{c\delta^{-k}}{(k+1)!}\binom{d}{k} \sum_{1\le i\le d+1} (d+1)\binom{d}{i-1} \\
&= \frac{c\delta^{-k}(d+1)}{(k+1)!}\binom{d}{k} 2^d \\
&= \frac{c\delta^{-k}}{k!}\binom{d+1}{k+1} 2^d \\
&\le c\delta^{-k}(4/k)^k 2^d 2^d
\end{aligned}$$

$$\leq c\left(\frac{4(d+1)}{k(m-d)}\right)^k 4^d$$

∎

The bound obtained above is not good for small k, but it is very good for large k. In fact a_k is so small for large k, that most terms of $p(x)$ can be practically ignored when x is small. We exploit this fact to prove that all the coefficients are small.

Lemma 10. *Let*

$$p(x) = \sum_{0\leq k\leq d} a_k x^k$$

be a dth degree polynomial, and let $c = \max_{1\leq i\leq m} |p(i)|$, *for some* $m \geq d+1$. *Then*

$$|a_1| \leq \frac{4cd^{10}}{m + d/255 + 1}.$$

Proof: The proof is by induction on d. When $d = 1$, $p(x)$ is a line, whose slope is obviously between $-2c/m$ and $2c/m$, so

$$|a_1| \leq 2c/m \leq \frac{2c}{m + 1/255 + 1}.$$

so the base case is established. Now assume that the result has been established for all degrees less than d. We prove it for d.

Let $k \geq \lceil (d+1)/2 \rceil$ and let $x \leq (m-d)/512$. By the preceding lemma,

$$|a_k| \leq c\left(\frac{4(d+1)}{k(m-d)}\right)^k 4^d \leq c(8/(m-d))^k 4^d$$

so

$$|a_k x^k| \leq c(8x/(m-d))^k 4^d \leq c(1/64)^k 4^d \leq c(1/8)^d 4^d = c2^{-d} \leq c/d.$$

Therefore, for $x \leq (m-d)/512$,

$$\left|\sum_{\lceil (d+1)/2\rceil \leq k\leq d} a_k x^k\right| \leq d(c/d) = c.$$

Let $q(x) = \sum_{0\leq k\leq d/2} a_k x^k$. Then, for $1 \leq i \leq \lfloor (m-d)/512 \rfloor$, $|q(i)| < p(i) + c \leq 2c$. Now, the degree of q is $\lceil (d+1)/2 \rceil - 1 \leq d/2$. In addition, the degree of q is at least 1 so we may apply the inductive hypothesis to bound a_1:

$$a_1 \leq \frac{4c(d/2)^{10}}{\lfloor (m-d)/512 \rfloor + d/510 + 1} \leq \frac{4c(d/2)^{10}}{(m-d-511)/512 + d/510 + 1} = \frac{4cd^{10}}{m + d/255 + 1},$$

completing the induction. ∎

Lemma 11. *Let*

$$p(x) = \sum_{0 \le k \le d} a_k x^k$$

be a dth degree polynomial, and let $c = \max_{1 \le i \le m} |p(i)|$, *for some* $m \ge d+1$. *Then, for* $k \ge 1$,

$$|a_k| \le 2\left(\frac{2}{k}\right)^k cd\left(\frac{d^9}{(m+d/255+1)}\right)^k.$$

Proof: When $k = 1$, this follows from Lemma 10, so assume that $k \ge 2$. The proof is by induction on d. For the base case, assume that $k \le d \le 2k-1$. By Lemma 9,

$$\begin{aligned}
|a_k| &\le c\left(\frac{4(d+1)}{k(m-d)}\right)^k 4^d \\
&\le c\left(\frac{8k}{k(m-d)}\right)^k 4^{2k-1} \qquad \text{because } d \le 2k-1 \\
&= \frac{1}{4}c\left(\frac{128}{(m-d)}\right)^k \\
&\le \frac{1}{4}c\left(\frac{d^7}{(m-d)}\right)^k \qquad \text{because } d \ge 2 \\
&\le \frac{1}{4}c\left(\frac{d^7(256d/255+2)}{(m+d/255+1)}\right)^k \qquad \text{because } m \ge d+1 \\
&= \frac{1}{4}c\left(\frac{256d/255+2}{d^2}\right)^k \left(\frac{d^9}{(m+d/255+1)}\right)^k \\
&\le \frac{1}{4}c\left(\frac{256k/255+2}{k^2}\right)^k \left(\frac{d^9}{(m+d/255+1)}\right)^k \qquad \text{because } d \ge k \\
&\le \frac{1}{4}c\left(\frac{2}{k}\right)^k \left(\frac{d^9}{(m+d/255+1)}\right)^k \qquad \text{because } k \ge 2 \\
&\le \frac{1}{8}\left(\frac{2}{k}\right)^k cd\left(\frac{d^9}{(m+d/255+1)}\right)^k \qquad \text{because } d \ge 2 \\
&\le 2\left(\frac{2}{k}\right)^k cd\left(\frac{d^9}{(m+d/255+1)}\right)^k,
\end{aligned}$$

establishing the base case. Now assume that the result has been established for all degrees less than some $d > 2k-1$.

Let $j \ge \lceil (d+1)/2 \rceil$ and let $x \le (m-d)/512$. As in the proof of Lemma 10, we have, for $x \le (m-d)/512$,

$$\left| \sum_{\lceil (d+1)/2 \rceil \le j \le d} a_j x^j \right| \le c.$$

Let $q(x) = \sum_{0 \le j \le d/2} a_j x^j$. Then, for $1 \le i \le \lfloor (m-d)/512 \rfloor$, $|q(i)| < p(i) + c \le 2c$. Now, the degree of q is $\lceil (d+1)/2 \rceil - 1 \le d/2$. In addition, the degree of q is at least k so we may apply the inductive hypothesis to bound a_k, as in the proof of Lemma 10:

$$\begin{aligned} a_k &\le 2\left(\frac{2}{k}\right)^k 2c(d/2)\left(\frac{(d/2)^9}{\lfloor (m-d)/512 \rfloor + d/510 + 1}\right)^k \\ &\le 2\left(\frac{2}{k}\right)^k cd\left(\frac{d^9}{m + d/255 + 1}\right)^k, \end{aligned}$$

completing the induction. ∎