HOW MANY ZEROS OF A RANDOM *SPARSE* POLYNOMIAL ARE REAL?

GORAV JINDAL, ANURAG PANDEY, HIMANSHU SHUKLA, AND CHARILAOS ZISOPOULOS

Abstract. We investigate the number of real zeros of a univariate k-sparse polynomial f over the reals, when the coefficients of f come from independent standard normal distributions. Recently Bürgisser, Ergür and Tonelli-Cueto showed that the expected number of real zeros of f in such cases is bounded by $O(\sqrt{k}\log k)$. In this work, we improve the bound to $O(\sqrt{k})$ and also show that this bound is tight by constructing a family of sparse support whose expected number of real zeros is lower bounded by $\Omega(\sqrt{k})$. Our main technique is an alternative formulation of the Kac integral by Edelman-Kostlan which allows us to bound the expected number of zeros of f in terms of the expected number of zeros of polynomials of lower sparsity. Using our technique, we also recover the $O(\log n)$ bound on the expected number of real zeros of a dense polynomial of degree n with coefficients coming from independent standard normal distributions.

1. Introduction

Understanding the number of real zeros of a given real univariate polynomial has always been of interest, both from a theoretical as well as an application point of view in science, engineering and mathematics. The arithmetic of sparse polynomials has been of special interest in computer science and the algorithms for efficiently finding roots of sparse polynomials have been widely studied (see [5, 19, 24, 9, 25, 26]).

1.1. Zeros of random sparse univariate polynomials

In order to gain a better understanding of the behavior of the number of real zeros for sparse polynomials and its generalizations, we study the case of a single univariate sparse random polynomial. For simplicity, in this article, we only consider the case when the coefficients are identically distributed independent standard normal random variables.

With respect to this consideration, the dense case[1] has been extensively studied and is well understood. This problem was first considered in a series of works by Littlewood and Offord [20, 21] who proved a $O(\log^2(n))$ upper bound on the expected number of real zeros in the dense case when the coefficients are from bernoulli($\{-1,1\}$),

2010 *Mathematics Subject Classification.* 68Q15, 34F05, 11C08, 30C15, 26D99.

Key words and phrases. Algebraic complexity theory, sparse polynomials, real-tau conjecture, random polynomials, real zeros.

[1] i.e. there is no restriction on the sparsity, thus we have a polynomial f of degree n with all its $n+1$ coefficients as standard normal random variables

standard gaussian and uniform distributions. In 1943, Kac [14] proved that the expected number of real zeros for degree n polynomial with coefficients drawn form standard normal distribution is:

$$\left(\frac{2}{\pi} + o(1)\right)\log(n).$$

Later, in a series of works, Offord, Erdős, Stevens, Ibragimov and Maslova [11, 30, 13] extended the results to other more general distributions including but not restricted to bernoulli ({-1,1}) and uniform distributions. Interested readers can look at the article by Erdélyi [10] for more recent results in this direction. In 1995, Edelman and Kostlan [8] gave an alternative, simpler derivation for the Kac bound using geometric methods, in addition to providing essential insights to the integral and numerous generalizations in a variety of cases. For this article, the works in [14, 8] are most relevant. It seems very surprising that there are so few real zeros in the random case.

In the sparse case, there is a line of work considering the case of the multivariate system of random equations (for instance see [15, 23, 22]). However their focus is different and we are not aware of any useful adaptations to the univariate case. In fact, we do not know of any such progress until the recent work of Bürgisser, Ergür and Tonelli-Cueto [4] which showed that for a random k-sparse univariate polynomial, the expected number of real roots in the standard normal case, is bounded by $\frac{4}{\pi}\sqrt{k}\log k$,[2] thus showing that in this setting, the number of real zeros is much less than the Descartes' bound.

1.2. Zeros of sparse polynomials

A lot of the polynomials that we encounter in practice are *sparse*, i.e. the number of monomials in them is considerably smaller than the degree of the polynomial. This motivates one to study the question for the sparse polynomials. Descartes' famous rule of signs from the 17th century [7] already sheds some light by bounding the number of non-zero real zeros of a k-sparse $f \in \mathbb{R}[x]$ [3] by $2k-2$. There are polynomials which achieve this bound too. Having some understanding on the number of real roots of k-sparse polynomials, it makes sense to ask the same question for generalizations. However, if we consider the first non-trivial generalization, i.e. if we consider the polynomial $fg+1$, where f and g are both k-sparse, our understanding seems very limited. To the best of our knowledge, no bound better than the one given by Descartes' rule of sign is known in this case, in particular, no sub-quadratic bound is known. We also do not know of any example where the number of real roots of $fg+1$ is super-linear in k.

1.3. Connections to algebraic complexity theory: Real Tau Conjecture

A strong motivation for computer scientists to consider generalizations like the above was provided in 2011 by Koiran [16], when he considered the number of real zeros of the sum of products of sparse polynomials. He formulated *the real τ-conjecture*

[2] unless stated otherwise, the base of the logarithm in this article is always e

[3] throughout this article, polynomials considered are over reals and have degree n with $n >> k$.

claiming that if a polynomial is given as

$$f = \sum_{i=1}^{m} \prod_{j=1}^{t} f_{ij}$$

where all f_{ij}'s are k-sparse, then the number of real zeros of f is bounded by a polynomial in mkt. Thus the conjecture claims that a univariate polynomial computed by a depth-4 arithmetic circuit (see [28, 27] for background on arithmetic circuits) with the fan-in of gates at the top three layers being bounded by m, t and k respectively will have $O((mkt)^c)$ real zeros for some positive constant c. Notice that applying Descartes' bound only gives an exponential bound on the number of real zeros of f, since a-priori the sparsity bound that we can achieve for f is only $O(mk^t)$.

What is of particular interest is the underlying connection of this conjecture to the central question of algebraic complexity theory. Koiran showed that proving the conjecture implies a superpolynomial lower bound on the arithmetic circuit complexity of the permanent, hence establishing the importance of the question of understanding real roots of sparse polynomials from the perspective of theory of computation as well. In fact this connection is what inspired the authors to investigate the problems considered in this article.

The real τ-conjecture itself was inspired by the Shub and Smale's τ-conjecture [29] which asserts that the number of integer zeros of a polynomial with arithmetic circuit complexity bounded by s will be bounded by a polynomial in s. This conjecture also implies a super-polynomial lower bound on the arithmetic circuit size of the permanent [3] and also implies $\mathrm{P}_{\mathbb{C}} \neq \mathrm{NP}_{\mathbb{C}}$ in the Blum-Shub-Smale model of computation (see [29, 1]). Koiran's motivation was to connect the complexity theoretic lower bounds to the number of real zeros instead of the number of integer zeros, because the latter takes one to the realm of number theory where problems become notoriously hard very easily.

While the real τ-conjecture remains open (see [12, 18, 17] for some works towards it), Briquel and Bürgisser [2] showed that the conjecture is true in the average case, i.e. they show that when the coefficients involved in the description of f are independent Gaussian random variables, then the expected number of real zeros of f is bounded by $O(mkt)$.

1.4. **Our contributions**

Before we state our results we set up some notations. Consider a set $S = \{e_1, \ldots, e_k\} \subseteq \mathbb{N}$ of natural numbers. For such a set S, one asks how many roots (in expectation) of the random polynomial $f_S = \sum_{i=1}^{k} a_i x^{e_i}$ (here a_i's are independent standard normals) are real. For an open interval $I \subseteq \mathbb{R}$, we use z_S^I to denote the expected number of roots of f_S in I. To avoid some degeneracy issues, we always assume $0 \notin I$, this assumption allows us to assume that the smallest element of S is zero. In this paper, we are only concerned with the case when $I = (0, 1)$. See Remark 1.2 on why this is sufficient. When $I = (0, 1)$, we simply use z_S to denote z_S^I.

Our main contribution is the improvement on the bound on the expected number of real zeros of a random k-sparse polynomial f and proving that this is the best one can do.

Theorem 1.1. *Let $S \subseteq \mathbb{N}$ be any set as above with $|S| = k$, then we have $z_S \leq \frac{2}{\pi}\sqrt{k-1}$.*

Remark 1.2. Since our bound in Theorem 1.1 only depends on the size of S, and not on the structure of S, we get that $z_S^{\mathbb{R}} = 4z_S^{(0,1)}$. For $S = \{e_1, \ldots, e_k\}$, $z_S^{(1,\infty)}$ is equal to $z_{S'}^{(0,1)}$ for $S' = \{n-e_1, \ldots, n-e_k\}$ by replacing x by $\frac{1}{x}$ and multiplying by x^n, where n is the degree of f_S. Also $z_S^{(-\infty,0)} = z_S^{(0,\infty)}$ by replacing x by $-x$.

Theorem 1.3. *There exists a sequence of sets $S_k \subset \mathbb{N}$ for $k \geq 1$ with $|S_k| = k+2$ and a constant $c > 0$ such that $z_{S_k} \geq c \cdot \sqrt{k}$, for large enough k.*

Theorem 1.3 shows that the bound obtained in Theorem 1.1 is tight and cannot be reduced further for an arbitrary $S \subset \mathbb{N}$.

Using our techniques, we confirm the intuition from the dense case that in expectation, all the roots are concentrated around 1 i.e. for any small constant $\epsilon > 0$, the expected number of roots in $(0, 1-\epsilon)$ is bounded by a constant independent of n and k.

Theorem 1.4. *For a fixed $\epsilon > 0$ and any $S \subseteq \mathbb{N}$ as above, we have*

$$z_S^{(0,1-\epsilon)} \leq \frac{1}{2\pi}\left(\log\left(\frac{2}{\epsilon}\right) + \frac{4}{\sqrt{\epsilon}} - 4\right).$$

1.5. **Proof ideas**

Our main technical contribution is an alternative formulation of the Kac integral by Edelman-Kostlan, that we call the *Edelman-Kostlan integral* (discussed in Section 2) presented in detail in Section 3.

The formulation allows us to bound $z_{S_1 \sqcup S_2}$ in terms of the bounds on z_{S_1} and z_{S_2} (presented in subsection 3.2). Thus we can build our k-sparse polynomial monomial-by-monomial. We show that every time we add a monomial, we do not increase the expected number of roots by a lot. A careful application of this idea yields the desired $O(\sqrt{k})$ bound (presented in Section 4).

We also obtain a bound on $z_{S_1+S_2}$ in terms of z_{S_1} and z_{S_2}, where $S_1 + S_2$ is the set obtained as a result of the addition of elements of S_1 and S_2 (presented in subsection 3.1). Combining the bounds on $z_{S_1+S_2}$ and $z_{S_1 \sqcup S_2}$ allows us to recover the $O(\log n)$ bound for the dense case i.e. $S = \{0, 1, \ldots, n\}$, where we build up our set S as a combination of unions and additions of sets (presented in Section 8.1).

Further, the proof that all the roots are concentrated around 1 follows from the analysis of an approximation of the Edelman-Kostlan integral. This approximation which is inspired by the one used in [4] makes the analysis of the integral simpler.

Finally in Section 6, we show that we cannot obtain a better bound for an arbitrary $S \subset \mathbb{N}$. We show this by applying the idea of monomial-wise construction of a polynomial (presented in Section 3.2) on a carefully chosen monomial sequence, thus proving Theorem 1.3.

1.6. **Previous work: known bounds on z_S^I**

In this subsection, we present the state of the art prior to this work for z_S^I.

For $S = \{0, 1, 2, \ldots, n\}$ and $I = \mathbb{R}$, z_S^I is known to be bounded by $O(\log n)$.

Theorem 1.5 ([8, 14]). *If $S = \{0, 1, 2, \dots, n\}$ then*

$$z_S^{\mathbb{R}} = \frac{2}{\pi}\log(n) + C_1 + \frac{2}{n\pi} + O\left(\frac{1}{n^2}\right).$$

Here $C_1 \approx 0.6257358072\dots$.

Determining the value of z_S^I for arbitrary sets S remains an open problem. Towards this the best bound known was the following result by [4].

Theorem 1.6 ([4, Theorem 1.3]). *Let $S \subseteq \mathbb{N}$ be any set as above with $|S| = k$ then we have*

$$z_S \le \frac{1}{\pi}\sqrt{k}\log(k).$$

For the sake of completeness, we present a proof for the above theorem in the appendix.

2. Preliminaries

Since our method builds upon the Edelman-Kostlan method [8] by a novel approach on analyzing their integral, it is essential to look at their method. In order to compute z_S for $S = \{e_1, \dots, e_k\}$, define a generalization of the moment curve v_S as $v_S(t) := (t^{e_1}, t^{e_2}, \dots, t^{e_k})$. This allows the following expression for z_S^I:

Theorem 2.1 ([8, Theorem 3.1]). *For all sets $S \subseteq \mathbb{N}$, we have the following equality for z_S^I .*

$$z_S^I = \frac{1}{\pi}\int_I \frac{\sqrt{(\|v_S(t)\|_2 \cdot \|v_S'(t)\|_2)^2 - (v_S(t)\cdot v_S'(t))^2}}{(\|v_S(t)\|_2)^2}\mathrm{d}t. \tag{1}$$

We refer to the above integral as the *Edelman-Kostlan integral*.

The strength of this method is that the above integral is parameterized by the support S and the interval I, thus allowing one to estimate the expected number of real zeros for any such arbitrary support and interval. In their paper, they compute the integral for $S = \{0, 1, \dots, k\}$ and $I = (0, 1)$ and for these values showed that z_S^I is bounded by $O(\log k)$. However, for arbitrary S of cardinality k, the integral becomes quite complicated to analyze.

In [4], they get around this difficulty by upper bounding the integral. This is achieved by ignoring the negative term of the numerator and through some elementary norm inequalities leads to the $O(\sqrt{k}\log k)$ bound.

We now state a basic technical proposition which will be useful in the proof of the main theorem.

Proposition 2.2. *The following identity is true for all a, b, c, d:*

$$\left(\frac{a+c}{b+d}\right)^2 = \left(\frac{b}{b+d}\right)\left(\frac{a}{b}\right)^2 + \left(\frac{d}{b+d}\right)\left(\frac{c}{d}\right)^2 - \frac{1}{bd}\left(\frac{bc-ad}{b+d}\right)^2.$$

Proof. The RHS of above equation can we written as:

$$\begin{aligned}\left(\frac{b}{b+d}\right)\left(\frac{a}{b}\right)^2+\left(\frac{d}{b+d}\right)\left(\frac{c}{d}\right)^2-\frac{1}{bd}\left(\frac{bc-ad}{b+d}\right)^2 &= \frac{a^2d(b+d)+c^2b(b+d)-(bc-ad)^2}{(b+d)^2bd}\\ &= \frac{bd(a^2+c^2)+a^2d^2+c^2b^2-(bc-ad)^2}{(b+d)^2bd}\\ &= \frac{bd(a^2+c^2+2ac)}{(b+d)^2bd}\\ &= \left(\frac{a+c}{b+d}\right)^2.\end{aligned}$$

□

3. Alternative formulation of the Edelman-Kostlan Integral

In this section, we present an alternative formulation of the Edelman-Kostlan integral on which our proofs build upon.

Definition 3.1. *For a set $S=\{e_1,e_2,\ldots,e_k\}\subseteq\mathbb{N}$, we define:*

$$g_S(t):=(\|v_S(t)\|_2)^2=\sum_{i=1}^{k}t^{2e_i}.$$

In the following lemma, we show that we can express z_S^I entirely in terms of $g_S(t)$ and its derivatives. Hence we define:

Definition 3.2. *For a function $g:\mathbb{R}\to\mathbb{R}$, we define the function $\mathcal{I}(g):\mathbb{R}\to\mathbb{R}$,*

$$\mathcal{I}(g):=\left(\frac{g'(t)}{g(t)}\right)'+\frac{g'(t)}{tg(t)}=(\log(g(t)))''+\frac{(\log(g(t)))'}{t}.$$

We are now ready give our alternative formulation.

Lemma 3.3. *For all sets $S\subseteq\mathbb{N}$, we have the following equality for z_S^I .*

$$z_S^I=\frac{1}{2\pi}\int_I\sqrt{\mathcal{I}(g_S(t))}\mathrm{d}t.$$

Proof. We can rewrite equation (1) as:

$$z_S^I=\frac{1}{\pi}\int_I\frac{\sqrt{(g_S(t)\cdot(\|v_S'(t)\|_2)^2-(v_S(t)\cdot v_S'(t))^2}}{g_S(t)}\mathrm{d}t.$$

Now verify the following equality for $v_S(t)\cdot(v_S'(t))$.

$$v_S(t)\cdot v_S'(t)=\sum_{i=1}^{k}e_it^{2e_i-1}=\frac{g_S'(t)}{2}$$

We also have the following equality for $(\|v_S'(t)\|_2)^2$.

$$\begin{aligned}
(\|v_S'(t)\|_2)^2 &= \sum_{i=1}^{k} e_i^2 t^{2e_i-2} = \frac{1}{4}\cdot\left(\sum_{i=1}^{k} 4e_i^2 t^{2e_i-2}\right) \\
&= \frac{1}{4}\cdot\left(\sum_{i=1}^{k}((2e_i(2e_i-1))+2e_i)\cdot t^{2e_i-2}\right) \\
&= \frac{1}{4}\cdot\left(\sum_{i=1}^{k}(2e_i(2e_i-1)\cdot t^{2e_i-2}\right)+\frac{1}{4}\cdot\left(\sum_{i=1}^{k}2e_i\cdot t^{2e_i-2}\right) \\
&= \frac{1}{4}g_S''(t)+\frac{1}{4t}g_S'(t).
\end{aligned}$$

Therefore we can rewrite z_S^I as:

$$\begin{aligned}
z_S^I &= \frac{1}{\pi}\int\limits_I \sqrt{\frac{1}{4}\left(\frac{g_S(t)\cdot(g_S''(t)+\frac{1}{t}g_S'(t))-(g_S'(t))^2}{(g_S(t))^2}\right)}\mathrm{d}t \\
&= \frac{1}{2\pi}\int\limits_I \sqrt{\frac{g_S''(t)}{g_S(t)}-\left(\frac{g_S'(t)}{g_S(t)}\right)^2+\frac{g_S'(t)}{tg_S(t)}}\mathrm{d}t \\
&= \frac{1}{2\pi}\int\limits_I \sqrt{\left(\frac{g_S'(t)}{g_S(t)}\right)'+\frac{g_S'(t)}{tg_S(t)}}\mathrm{d}t \\
&= \frac{1}{2\pi}\int\limits_I \sqrt{(\log(g_S(t)))''+\frac{(\log(g_S(t)))'}{t}}\mathrm{d}t.
\end{aligned}$$

□

As can be seen from the above, whenever the Edelman-Kostlan integral is well defined, the conditions on g which make $\mathcal{I}(g)$ well defined and non-negativity conditions are also satisfied. This is true for all cases we consider, i.e. for every S, $g_S(t)$ satisfies all the needed conditions.

We now give a useful characterization of $\mathcal{I}(g_S(t))$ that will be used to show that in a very small interval I near 1, z_S^I is very small (see Lemma 6.1 and Remark 6.2).

Proposition 3.4. *For $S=\{e_1,e_2,\dots,e_k\}\subseteq\mathbb{N}$, $\mathcal{I}(g_S(t))$ satisfies the following equality:*

$$\mathcal{I}(g_S(t)) = \frac{4}{(g_S(t))^2}\cdot\left(\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} ((e_i-e_j)t^c)^2\right).$$

Proof. We have:

$$\begin{aligned}(g_S(t)\cdot(\|v_S'(t)\|_2)^2-(v_S(t)\cdot v_S'(t))^2 &= \left(\sum_{i=1}^{k} t^{2e_i}\right)\cdot\left(\sum_{j=1}^{k} e_j^2 t^{2e_j-2}\right)-\left(\sum_{i=1}^{k} e_i t^{2e_i-1}\right)^2\\ &= \sum_c \sum_{\substack{i,j\\ e_i+e_j-1=c}} e_j^2 t^{2c}-\sum_c \sum_{\substack{i,j\\ e_i+e_j-1=c}} e_i\cdot e_j t^{2c}\end{aligned}$$

$$\begin{aligned}&=\left(\sum_c \sum_{\substack{i\\ e_i+e_i-1=c}} e_i^2 t^{2c}+\sum_c \sum_{\substack{i\neq j\\ e_i+e_j-1=c}} e_j^2 t^{2c}\right)-\left(\sum_c \sum_{\substack{i\\ e_i+e_i-1=c}} e_i^2 t^{2c}+\sum_c \sum_{\substack{i\neq j\\ e_i+e_j-1=c}} e_i e_j t^{2c}\right)\\ &=\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} (e_i^2+e_j^2-2e_ie_j)t^{2c}=\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} ((e_i-e_j)t^c)^2.\end{aligned}$$

□

Remark 3.5. Using the above proposition and Lemma 3.3 we have:

$$z_S^I=\frac{1}{\pi}\int_I \frac{\sqrt{\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} ((e_i-e_j)t^c)^2.}}{g_S(t)}\mathrm{d}t. \tag{2}$$

The formulation in Definition 3.2 allows us to prove the following proposition:

Proposition 3.6. *For two non-negative functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}$, we have that $\sqrt{\mathcal{I}(g_1\cdot g_2)} \le \sqrt{\mathcal{I}(g_1)}+\sqrt{\mathcal{I}(g_2)}$.*

Proof. Consider:

$$\mathcal{I}(g_1\cdot g_2)=(\log(g_1(t)\cdot g_2(t)))''+\frac{(\log(g_1(t)\cdot g_2(t)))'}{t}$$

(By linearity of differentiation and the fact that $\log(g_1\cdot g_2)=\log(g_1)+\log(g_2)$)

$$\begin{aligned}&=(\log(g_1(t)))''+\frac{(\log(g_1(t)))'}{t}+(\log(g_2(t)))''+\frac{(\log(g_2(t)))'}{t}\\ &=\mathcal{I}(g_1)+\mathcal{I}(g_2).\end{aligned}$$

Now the claim follows by using the fact that $\sqrt{x+y}\le\sqrt{x}+\sqrt{y}$ for non-negative x, y. □

This allows us to give a bound on the integral when $S = S_1 * S_2$, where $*$ corresponds to the operation of either union or addition of sets. This bound depends on the integrals associated to the corresponding sets S_1 and S_2.

3.1. **Addition of sets**

Definition 3.7. *For sets $A, B \subseteq \mathbb{N}$, we define the sum of A, B as: $A+B := \{a+b : a \in A, b \in B\}$. We say two sets $A, B \subseteq \mathbb{N}$ are* collision-free *if $|A+B| = |A|\cdot|B| = |A\times B|$, i.e. when all the "$a+b : a\in A, b\in B$" are distinct.*

Now we show how to apply this definition in the context of the above formulation of z_S^I and $\mathcal{I}(g)$.

Lemma 3.8. *If $S_1, S_2 \subseteq \mathbb{N}$ are two collision-free sets (as defined in Definition 3.7), then $z^I_{S_1+S_2} \le z^I_{S_1} + z^I_{S_2}$.*

Proof. It is easy to see from the definition of g_S, when S_1, S_2 are collision-free, we have:

$$g_{S_1+S_2}(t) = g_{S_1}(t) \cdot g_{S_2}(t).$$

Therefore we obtain:

$$\begin{aligned}
z^I_{S_1+S_2} &= \frac{1}{2\pi}\int_I \sqrt{\mathcal{I}(g_{S_1+S_2}(t))}\mathrm{d}t = \frac{1}{2\pi}\int_I \sqrt{\mathcal{I}(g_{S_1}(t)\cdot g_{S_2}(t))}\mathrm{d}t \\
\text{(Follows from 3.6)} \qquad &\le \frac{1}{2\pi}\int_I \sqrt{\mathcal{I}(g_{S_1}(t))}\mathrm{d}t + \frac{1}{2\pi}\int_I \sqrt{\mathcal{I}(g_{S_2}(t))}\mathrm{d}t \\
&= z^I_{S_1} + z^I_{S_2}.
\end{aligned}$$

□

Corollary 3.9. *If $S = \{0,1,\dots,k\} \cup \{n-k, n-k+1, \dots, n\}$ with $n > 2k$, then $z_{S_1+S_2} \le O(\log k)$.*

Proof. Note that $S = \{0,1,\dots,k\} + \{0, n-k\}$. Now the result follows by using Theorems 1.5 and 1.6, and Lemma 3.8. □

3.2. Union of sets

In subsection 3.1, we demonstrated an upper bound $z_{S_1+S_2}$ in terms of z_{S_1} and z_{S_2}. In this section, we want to find upper bounds for $z_{S_1 \sqcup S_2}$, here $S_1 \sqcup S_2$ denotes the disjoint union of S_1 and S_2. First we state the following proposition which is easy to verify.

Proposition 3.10. *If $S_1, S_2 \subseteq \mathbb{N}$ are two disjoint sets then $g_{S_1 \sqcup S_2}(t) = g_{S_1}(t) + g_{S_2}(t)$.*

We need the following definition to give our result for expressing $z_{S_1 \sqcup S_2}$ in terms of z_{S_1} and z_{S_2}.

Definition 3.11. *If $S_1, S_2 \subseteq \mathbb{N}$ are two disjoint sets with $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ being non-negative at zero.[4] Let $c_1, \dots, c_m$ (with $c_i \le c_{i+1}$) be the critical points of odd multiplicity of $\frac{g_{S_1}}{g_{S_2}}$ in $(0,1)$. Define $c_0 := 0$ and $c_{m+1} := 1$. We define the following quantities, here $0 \le i \le m$ and $c \in (0,1)$.*

[4]Note that at least one of $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ and $\left(\frac{g_{S_2}}{g_{S_1}}\right)'$ has to be non-negative at zero. Thus, we can always rename accordingly S_1 and S_2 to ensure this is the case.

$$\gamma_{S_1,S_2}(c) = \sqrt{\frac{g_{s_1}(c)}{g_{s_2}(c)}}$$

$$T^i_{S_1,S_2} := \begin{cases} \arctan(\gamma_{S_1,S_2}(c_{i+1})) - \arctan(\gamma_{S_1,S_2}(c_i)) & \textit{If } i \textit{ is even} \\ \arctan\left(\frac{1}{\gamma_{S_1,S_2}(c_{i+1})}\right) - \arctan\left(\frac{1}{\gamma_{S_1,S_2}(c_i)}\right) & \textit{If } i \textit{ is odd} \end{cases}$$

$$R_{S_1,S_2} := \sum_{i=0}^{m} T^i_{S_1,S_2}.$$

Lemma 3.12. *Let $S_1, S_2 \subseteq \mathbb{N}$ be two disjoint sets. WLOG assume that $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is non-negative at zero. Then we have:*

$$z_{S_1 \sqcup S_2} \leq z_{S_1} + z_{S_2} + \frac{1}{\pi} R_{S_1,S_2}.$$

Proof. By using Proposition 3.10, we know that:

$$\mathcal{I}(g_{S_1 \sqcup S_2}) = \mathcal{I}(g_{S_1} + g_{S_2}) = \frac{g''_{S_1} + g''_{S_2}}{g_{S_1} + g_{S_2}} - \left(\frac{g'_{S_1} + g'_{S_2}}{g_{S_1} + g_{S_2}}\right)^2 + \frac{1}{t}\left(\frac{g'_{S_1} + g'_{S_2}}{g_{S_1} + g_{S_2}}\right)$$

(Follows by applying Proposition 2.2 on $g'_{S_1} = a, g_{S_1} = b, g'_{S_2} = c, g_{S_2} = d$)

$$= \frac{g_{S_1}}{g_{S_1} + g_{S_2}} \cdot \mathcal{I}(g_{S_1}) + \frac{g_{S_2}}{g_{S_1} + g_{S_2}} \cdot \mathcal{I}(g_{S_2}) + \frac{1}{g_{S_1} g_{S_2}} \left(\frac{g_{S_1} g'_{S_2} - g_{S_2} g'_{S_1}}{g_{S_1} + g_{S_2}}\right)^2.$$

Therefore we have:

$$\begin{aligned} z_{S_1 \sqcup S_2} &= \frac{1}{2\pi} \int_0^1 \sqrt{\mathcal{I}(g_{S_1 \sqcup S_2}(t))} \mathrm{d}t \\ &= \frac{1}{2\pi} \int_0^1 \sqrt{\frac{g_{S_1}}{g_{S_1} + g_{S_2}} \cdot \mathcal{I}(g_{S_1}) + \frac{g_{S_2}}{g_{S_1} + g_{S_2}} \cdot \mathcal{I}(g_{S_2}) + \frac{1}{g_{S_1} g_{S_2}} \cdot \frac{(g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2})^2}{(g_{S_1} + g_{S_2})^2}} \mathrm{d}t \\ &\leq \frac{1}{2\pi} \cdot \left(\int_0^1 \sqrt{\mathcal{I}(g_{S_1}(t))} \mathrm{d}t + \int_0^1 \sqrt{\mathcal{I}(g_{S_2}(t))} \mathrm{d}t + \int_0^1 \left| \frac{1}{\sqrt{g_{S_1} g_{S_2}}} \cdot \frac{g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}}{g_{S_1} + g_{S_2}} \right| \mathrm{d}t \right) \\ &= z_{S_1} + z_{S_2} + \frac{1}{2\pi} \int_0^1 \left| \frac{1}{\sqrt{g_{S_1} g_{S_2}}} \left(\frac{g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}}{g_{S_1} + g_{S_2}} \right) \right| \mathrm{d}t. \end{aligned}$$

Now we just need to upper bound the definite integral:

$$J := \int_0^1 \left| \frac{1}{\sqrt{g_{S_1} g_{S_2}}} \left(\frac{g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}}{g_{S_1} + g_{S_2}} \right) \right| \mathrm{d}t.$$

The value of J in a sub-interval (α, β) of $(0,1)$ depends upon the condition whether $g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}$ is positive or negative in (α, β). So we divide $(0,1)$ in the intervals where $g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}$ is positive or negative. Note that $g_{S_2} g'_{S_1} - g_{S_1} g'_{S_2}$ is positive if and only if $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is positive. Therefore $g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2}$ changes sign exactly on the critical points of odd multiplicity of $\frac{g_{S_1}}{g_{S_2}}$. Suppose (α, β) is some sub-interval of $(0,1)$ where $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is non-negative. Let us look at the integral J in the interval (α, β). We have:

$$J_{\alpha,\beta} := \int_{\alpha}^{\beta} \frac{1}{\sqrt{g_{S_1}g_{S_2}}} \left(\frac{g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2}}{g_{S_2}^2}\right) \cdot \left(\frac{g_{S_2}^2}{g_{S_1} + g_{S_2}}\right) \mathrm{d}t$$

Now we use the substitution $u = \sqrt{\frac{g_{s_1}}{g_{s_2}}}$ to obtain:

$$\begin{aligned} J_{\alpha,\beta} &:= \int_{\alpha}^{\beta} \frac{1}{\sqrt{g_{S_1}g_{S_2}}} \left(\frac{g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2}}{g_{S_2}^2}\right) \cdot \left(\frac{g_{S_2}^2}{g_{S_1} + g_{S_2}}\right) \mathrm{d}t \\ &= \int_{\alpha}^{\beta} \sqrt{\frac{g_{S_2}}{g_{S_1}}} \cdot \left(\frac{g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2}}{g_{S_2}^2}\right) \cdot \left(\frac{g_{S_2}}{g_{S_1} + g_{S_2}}\right) \mathrm{d}t \end{aligned}$$

(Here $\gamma = \sqrt{\frac{g_{s_1}(\alpha)}{g_{s_2}(\alpha)}}$ and η=$\sqrt{\frac{g_{s_1}(\beta)}{g_{s_2}(\beta)}}$.)

$$= 2\int_{\alpha}^{\beta} \left(\sqrt{\frac{g_{s_1}}{g_{s_2}}}\right)' \cdot \left(\frac{1}{1 + \left(\sqrt{\frac{g_{s_1}}{g_{s_2}}}\right)^2}\right) \mathrm{d}t = 2\int_{\gamma}^{\eta} \left(\frac{1}{1+u^2}\right) \mathrm{d}u$$

Therefore $J_{\alpha,\beta} = 2(\arctan(\eta) - \arctan(\gamma))$ with $\gamma = \sqrt{\frac{g_{s_1}(\alpha)}{g_{s_2}(\alpha)}}$ and $\eta = \sqrt{\frac{g_{s_1}(\beta)}{g_{s_2}(\beta)}}$. For intervals where $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is negative, we obtain the same result by using the substitution $u = \sqrt{\frac{g_{s_2}}{g_{s_1}}}$ instead, which is reflected on the definition of $T^i_{S_1,S_2}$ above. Now the claimed inequality for $z_{S_1 \sqcup S_2}$ follows by using the quantities defined in Definition 3.11. □

Corollary 3.13. *Let $S_1, S_2 \subseteq \mathbb{N}$ be two disjoint sets. WLOG assume that $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is non-negative at zero. If $\frac{g_{S_1}}{g_{S_2}}$ has m critical points in $(0,1)$ of odd multiplicity, then:*

$$z_{S_1 \sqcup S_2} \leq z_{S_1} + z_{S_2} + \frac{m+1}{2}.$$

Proof. By using Lemma 3.12, we know that:

$$z_{S_1 \sqcup S_2} \leq z_{S_1} + z_{S_2} + \frac{1}{\pi} R_{S_1,S_2}.$$

Also, $R_{S_1,S_2} := \sum_{i=0}^{m} T^i_{S_1,S_2}$. By using the definition of $T^i_{S_1,S_2}$ defined in Definition 3.11, it is clear that $T^i_{S_1,S_2} \leq \frac{\pi}{2}$. Therefore we have: $R_{S_1,S_2} \leq (m+1) \cdot \frac{\pi}{2}$. Hence the claimed inequality follows. □

4. Proof of Theorem 1.1: $O(\sqrt{k})$ bound

Proposition 4.1. *For any singleton set S, we have $\mathcal{I}(g_S) = 0$.*

Proof. Suppose $S = \{a\}$, therefore $g_S(t) = t^{2a}$. Hence:

$$\mathcal{I}(g_S) = (\log(g_S(t)))'' + \frac{(\log(g_S(t)))'}{t} = (2a\log(t))'' + \frac{(2a\log(t))'}{t}$$
$$= \frac{-2a}{t^2} + \frac{2a}{t^2} = 0.$$

□

Lemma 4.2. *For all sets S of size two, $z_S = \frac{1}{4}$.*

Proof. WLOG we can assume that $S = \{0, a\}$. We have:

$$z_S = \frac{1}{2\pi}\int_I \sqrt{\mathcal{I}(g_S(t))}\mathrm{d}t.$$

An easy calculation shows that $\sqrt{\mathcal{I}(g_S(t))} = \frac{2at^{a-1}}{1+t^{2a}}$. Therefore:

$$z_S = \frac{2}{2\pi}\int_0^1 \frac{at^{a-1}}{1+t^{2a}}\mathrm{d}t = \frac{1}{4}.$$

□

Now we show that if we increase the sparsity of a polynomial by adding a monomial of degree higher than the degree of the polynomial, we can bound the expected number of real zeros of the resulting polynomial in terms of the bound on the expected number of zeros of the original polynomial.

Lemma 4.3. *Let $S \subseteq \mathbb{N}$ be a set with $0 \in S$ and $|S| = k$. If $a \in \mathbb{N}$ is such that $a > \max(S)$ then:*

$$z_{S\cup\{a\}} \leq z_S + \frac{1}{\pi}\arctan\left(\frac{1}{\sqrt{k}}\right).$$

Proof. Let us first analyze the derivative of $\frac{g_{\{a\}}}{g_S}$.We have:

$$\left(\frac{g_{\{a\}}}{g_S}\right)' = \frac{\left(g_{\{a\}}\right)' g_S - \left(g_S\right)' g_{\{a\}}}{g_S^2} = \frac{1}{g_S^2}\left(2ax^{2a-1}\sum_{e\in S} x^{2e} - x^{2a}\sum_{e\in S} 2ex^{2e-1}\right) \tag{3}$$

$$= \frac{2x^{2a-1}}{g_S^2}\left(\sum_{e\in S}(a-e)x^{2e}\right) > 0. \tag{4}$$

Therefore $\frac{g_{\{a\}}}{g_S}$ is always increasing in $(0,1)$. Hence we have:

$$\begin{aligned}
z_{S\cup\{a\}} &= \frac{1}{2\pi}\int_0^1 \sqrt{\mathcal{I}(g_{S\cup\{a\}}(t))}\mathrm{d}t \\
&= \frac{1}{2\pi}\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)+\frac{g_{\{a\}}}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_{\{a\}})+\frac{1}{g_S g_{\{a\}}}\left(\frac{\left(g_{\{a\}}\right)' g_S-\left(g_S\right)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)^2}\mathrm{d}t \\
&\le \frac{1}{2\pi}\cdot\left(\int_0^1 \sqrt{\mathcal{I}(g_S(t))}\mathrm{d}t+0+\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{\left(g_{\{a\}}\right)' g_S-\left(g_S\right)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)\mathrm{d}t\right) \\
&= z_S+\frac{1}{2\pi}\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{\left(g_{\{a\}}\right)' g_S-\left(g_S\right)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)\mathrm{d}t.
\end{aligned}$$

Now we use the substitution $u=\sqrt{\frac{g_{\{a\}}}{g_s}}$ to obtain:

(Here $\alpha=\sqrt{\frac{g_{\{a\}}(0)}{g_s(0)}}=0$ and $\beta{=}\sqrt{\frac{g_{\{a\}}(1)}{g_s(1)}}{=}\frac{1}{\sqrt{k}}$.)

$$\begin{aligned}
\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{\left(g_{\{a\}}\right)' g_S-\left(g_S\right)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)\mathrm{d}t &= 2\int_\alpha^\beta \left(\frac{1}{1+u^2}\right)\mathrm{d}u \\
&= 2\left(\arctan\left(\frac{1}{\sqrt{k}}\right)-\arctan(0)\right) \\
&= 2\arctan\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}$$

Hence:

$$z_{S\cup\{a\}} \le z_S+\frac{1}{2\pi}\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{\left(g_{\{a\}}\right)' g_S-\left(g_S\right)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)\mathrm{d}t = z_S+\frac{1}{\pi}\arctan\left(\frac{1}{\sqrt{k}}\right).$$

□

Theorem 4.4 (Theorem 1.1 restated)**.** *Let $S\subseteq\mathbb{N}$ be a set with $0\in S$ and $|S|=k$. Then $z_S\le\frac{1}{4}+\frac{2}{\pi}(\sqrt{k-1}-1)\le\frac{2}{\pi}\cdot\left(\sqrt{k}-1\right)$.*

Proof. If $k\le 2$ then the results follows from Lemma 4.2. So assume $k>2$. By using Lemma 4.2 and Lemma 4.3, we obtain that:

(We always add the highest element iteratively) $$z_S\le\frac{1}{4}+\frac{1}{\pi}\sum_{i=2}^{k-1}\arctan\left(\frac{1}{\sqrt{i}}\right).$$

We use the following well known inequality:

$$\arctan(x)<x \text{ for all } x>0.$$

This implies that:

$$z_S \le \frac{1}{4} + \frac{1}{\pi}\sum_{i=2}^{k-1}\frac{1}{\sqrt{i}}.$$

Now notice that:

$$\sum_{i=2}^{k-1}\frac{1}{\sqrt{i}} \le \int_{1}^{k-1}\sqrt{\frac{1}{x}}\mathrm{d}x = 2(\sqrt{k-1}-1).$$

Hence the claimed bound follows. □

5. Roots concentrate around 1: Proof of Theorem 1.4

Here we want to show that most of the roots are near 1. First we need the following proposition useful in the analysis.

Proposition 5.1. *For all $t \in (0,1)$, we have* $\sqrt{\sum_{e>0} e^2 t^{2e-2}} \le \frac{1}{1-t^2} + \frac{2t}{(1-t^2)^{\frac{3}{2}}}$.

Proof. First use the following well known equality:

$$\frac{1}{1-t^2} = \sum_{e\ge 0} t^{2e}.$$

Using this, we obtain that:

$$\left(\frac{1}{1-t^2}\right)'' = \sum_{e>0} 2e(2e-1)t^{2e-2} = \frac{2(1+3t^2)}{(1-t^2)^3}.$$

Therefore:

$$\sum_{e>0} e(2e-1)t^{2e-2} = \frac{(1+3t^2)}{(1-t^2)^3}.$$

Clearly:

$$\begin{aligned}\sqrt{\sum_{e>0} e^2 t^{2e-2}} \le \sqrt{\sum_{e>0} e(2e-1)t^{2e-2}} \le \sqrt{\frac{(1+3t^2)}{(1-t^2)^3}} = \sqrt{\frac{1}{(1-t^2)^2} + \frac{4t^2}{(1-t^2)^3}} \\ \le \frac{1}{1-t^2} + \frac{2t}{(1-t^2)^{\frac{3}{2}}}.\end{aligned}$$

□

We now give the proof of Theorem 1.4.

Proof. (Proof of Theorem 1.4). WLOG, we can assume that $0 \in S$, therefore $\|v_S(t)\|_2 \ge 1$ for all $t \in \mathbb{R}$. By using the equality in Theorem 2.1 and also by ignoring the second term in (1), we get the following inequality for z_S:

$$z_S^{(0,1-\epsilon)} \le \frac{1}{\pi}\int_0^{1-\epsilon} \frac{\sqrt{(\|v_S(t)\|_2 \cdot \|v'_S(t)\|_2)^2}}{(\|v_S(t)\|_2)^2}\mathrm{d}t = \frac{1}{\pi}\int_0^{1-\epsilon} \frac{\|v'_S(t)\|_2}{\|v_S(t)\|_2}\mathrm{d}t$$
$$\le \frac{1}{\pi}\int_0^{1-\epsilon} \|v'_S(t)\|_2 \,\mathrm{d}t$$

By using 5.1, we have: $\|v'_S(t)\|_2 = \sqrt{\sum_{e\in S} e^2 t^{2e-2}} \le \frac{1}{1-t^2} + \frac{2t}{(1-t^2)^{\frac{3}{2}}}$. Therefore:

$$z_S^{(0,1-\epsilon)} \le \frac{1}{\pi}\int_0^{1-\epsilon} \|v'_S(t)\|_2 \,\mathrm{d}t \le \frac{1}{\pi}\int_0^{1-\epsilon}\left(\frac{1}{1-t^2} + \frac{2t}{(1-t^2)^{\frac{3}{2}}}\right)\mathrm{d}t$$
$$= \frac{1}{\pi}\left(\int_0^{1-\epsilon}\frac{1}{1-t^2}\mathrm{d}t + \frac{1}{\pi}\int_0^{1-\epsilon}\frac{2t}{(1-t^2)^{\frac{3}{2}}}\mathrm{d}t\right)$$
$$= \frac{1}{\pi}\left(\left[\frac{1}{2}\log\left(\frac{1+t}{1-t}\right)\right]_0^{1-\epsilon} + \left[\frac{2}{\sqrt{1-t^2}}\right]_0^{1-\epsilon}\right)$$
$$= \frac{1}{\pi}\left(\frac{1}{2}\log\left(\frac{2-\epsilon}{\epsilon}\right) + \frac{2}{\sqrt{\epsilon(2-\epsilon)}} - 2\right) \le \frac{1}{2\pi}\left(\log\left(\frac{2}{\epsilon}\right) + \frac{4}{\sqrt{\epsilon}} - 4\right).$$

□

6. The lower bound

In this section we will come up with a sequence of sets $(S_k)_{k\ge1}$ such that the expected number of real zeros of the corresponding polynomials is lower bounded by $\Omega(\sqrt{k})$, for large enough k.

Lemma 6.1. *Suppose* $S = \{e_1, e_2, \dots, e_k\}$ *with* $e_k = \max(S)$ *and* $b \ge 1$. *If* $I_1 = \int_{1-\frac{1}{b}}^{1} \sqrt{\mathcal{I}(g_S)}\mathrm{d}t$, *then* $I_1 \le \frac{2(k+1)e_k}{b}$.

Proof. We have:

(By using Proposition 3.4)
$$\sqrt{\mathcal{I}(g_S)} = 2\frac{\sqrt{\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} (e_i-e_j)^2 t^{2c}}}{g_S}$$
$$\le 2\frac{\sqrt{\sum_c \sum_{\substack{i<j\\ e_i+e_j-1=c}} (e_{k+1})^2}}{g_S}$$
$$\le 2\frac{\sqrt{(k+1)^2 e_k^2}}{g_S}$$
$$\le 2(k+1)e_k.$$

Therefore, we have: $I_1 \le \int_{1-\frac{1}{b}}^{1} 2(k+1)e_k = \frac{2(k+1)e_k}{b}$. □

Remark 6.2. In the view of the above lemma, we can have z_S^I arbitrarily small, with $I = (1 - \frac{1}{b}, 1)$ for a large enough b. This fact will be crucial in the proof of Theorem 1.3. Further, Lemma 6.1 can be viewed as a supplementary result to Theorem 1.4. Theorem 1.4 implies that most of the roots lie in $(0, 1-\epsilon)$, if ϵ is allowed to be arbitrarily small. Lemma 6.1 gives a precise formulation of this fact.

From now on we will assume that $S = \{0,1\} \bigcup \{2^{2^i} \mid 1 \le i \le k-1\}$ and $a = 2^{2^k}$. The following lemma essentially will imply that one cannot avoid summing over $\sqrt{\frac{1}{k}}$ as in the proof of Theorem 4.4.

Lemma 6.3. *Let* $W = \frac{1}{g_S g_{\{a\}}}\left(\frac{(g_{\{a\}})' g_S - (g_S)' g_{\{a\}}}{g_S + g_{\{a\}}}\right)^2$, *then* $\int_{1-\frac{1}{2a}}^{1} \sqrt{W} \mathrm{d}t \ge \frac{c}{\sqrt{k}}$ *for some real constant* $c > 0$.

Proof. Using the computation in the proof of Lemma 4.3 we have:

$$\int_{1-\frac{1}{2a}}^{1} \sqrt{W} \mathrm{d}t = 2\left(\arctan\left(\frac{1}{\sqrt{k+1}}\right) - \arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right)\right).$$

We now upper bound the value of $\arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right)$ by giving a lower bound on $g_S(1-\frac{1}{2a})$ and an upper bound on $g_{\{a\}}(1-\frac{1}{2a})$. Using the well known inequalities $\left(1-\frac{1}{n}\right)^n \le \frac{1}{e}$ (for any $n \in \mathbb{N}$) and

$$(1+x)^r \ge 1 + rx \text{ if } x \ge -1 \text{ and } r > 1,$$

we have, for large enough k:

$$g_S(1-\frac{1}{2a}) = \sum_{i=1}^{k+1}\left(1-\frac{1}{2a}\right)^{2e_i} \ge \sum_{i=1}^{k+1}\left(1-\frac{2e_i}{2a}\right) \ge k+1-\left(\sum_{i=1}^{k+1} 2^{-k}\right) \ge k \tag{5}$$

Therefore,

$$\arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right) \le \arctan\left(\sqrt{\frac{\frac{1}{e}}{k}}\right)$$

$$\begin{aligned}
\Rightarrow 2\left(\arctan\left(\frac{1}{\sqrt{k+1}}\right) - \arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right)\right) &\ge 2\arctan\left(\frac{1}{\sqrt{k+1}}\right) \\
&\quad - 2\arctan\left(\sqrt{\frac{\frac{1}{e}}{k}}\right) \\
&\ge 2\left(\arctan\left(\frac{\frac{1}{\sqrt{k+1}} - \frac{1}{e\sqrt{k}}}{1+\frac{1}{e\sqrt{k(k+1)}}}\right)\right) \\
\text{(for some } c' > 0\text{)} \qquad &= 2\left(\arctan(\frac{c'}{\sqrt{k}})\right).
\end{aligned}$$

□

6.1. **Proof of Theorem 1.3**

For proving Theorem 1.3 we will again resort to our idea of monomial-wise construction of polynomial. The monomial sequence we choose is $e_{i+2} = 2^{2^i}$ for $i \geq 1$ with $e_1 = 0, e_2 = 1$. Before we begin the proof, recall from the proof of Lemma 4.3 that

$$z_{S\cup\{a\}} = \frac{1}{2\pi}\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + \frac{g_{\{a\}}}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_{\{a\}}) + \frac{1}{g_S g_{\{a\}}}\left(\frac{\left(g_{\{a\}}\right)' g_S - (g_S)' g_{\{a\}}}{g_S+g_{\{a\}}}\right)^2}\,\mathrm{d}t$$

(using the notation in Lemma 6.3)

$$= \frac{1}{2\pi}\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + \frac{g_{\{a\}}}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_{\{a\}}) + W}\,\mathrm{d}t.$$

The key idea is to write $z_{S\cup\{a\}}$ as a sum of two integrals over disjoint intervals such that $\mathcal{I}(g_S)$ dominates in one interval while W dominates in the other. The rest of the proof is about proving lower bounds on these two integrals.

Proof. We have:

$$\begin{aligned}
z_{S\cup\{a\}} &= \frac{1}{2\pi}\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + \frac{g_{\{a\}}}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_{\{a\}}) + W}\,\mathrm{d}t\\
&= \frac{1}{2\pi}\cdot\left(\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + 0 + W}\,\mathrm{d}t\right)\\
&= \frac{1}{2\pi}\left(\int_0^{1-\frac{1}{2a}} \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + W}\,\mathrm{d}t + \int_{1-\frac{1}{2a}}^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + W}\,\mathrm{d}t\right)\\
&\geq \frac{1}{2\pi}\left(\int_0^{1-\frac{1}{2a}} \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\,\mathrm{d}t + \int_{1-\frac{1}{2a}}^1 \sqrt{W}\,\mathrm{d}t\right)\\
&= \frac{1}{2\pi}\left(\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\,\mathrm{d}t - \int_{1-\frac{1}{2a}}^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\,\mathrm{d}t + \int_{1-\frac{1}{2a}}^1 \sqrt{W}\,\mathrm{d}t\right)
\end{aligned}$$

($\frac{g_{\{a\}}}{g_S}$ is increasing (Equation (3)))

$$\begin{aligned}
&\geq \frac{1}{2\pi}\left(\sqrt{\frac{k+1}{k+2}}\cdot\int_0^1 \sqrt{\mathcal{I}(g_S)}\,\mathrm{d}t - \int_{1-\frac{1}{2a}}^1 \sqrt{\mathcal{I}(g_S)}\,\mathrm{d}t + \int_{1-\frac{1}{2a}}^1 \sqrt{W}\,\mathrm{d}t\right)\\
&= \sqrt{\frac{k+1}{k+2}}\cdot z_S + \frac{1}{2\pi}\left(-\int_{1-\frac{1}{2a}}^1 \sqrt{\mathcal{I}(g_S)}\,\mathrm{d}t + \int_{1-\frac{1}{2a}}^1 \sqrt{W}\,\mathrm{d}t\right).
\end{aligned}$$

Now by using Lemma 6.1 with $b = 2a$ and Lemma 6.3 we have:

$$\begin{aligned} z_{S\cup\{a\}} &\geq \sqrt{\frac{k+1}{k+2}} \cdot z_S + \frac{1}{2\pi}\left(\int_{1-\frac{1}{2a}}^{1} \sqrt{W}\mathrm{d}t - I_1\right) \\ &\geq \sqrt{\frac{k+1}{k+2}} \cdot z_S + \frac{1}{\pi}\arctan\left(\frac{c'}{\sqrt{k}}\right) - \frac{k+1}{2\pi 2^{2^{k-1}}}. \end{aligned}$$

By a generalization of the Shafer-Fink inequality [6, Theorem 1], we have $z_{S\cup\{a\}} \geq \sqrt{\frac{k+1}{k+2}} \cdot z_S + \frac{c''}{\sqrt{k}}$ for some $c'' \in (0,1)$. Therefore, for large enough k, and i such that $k-i$ is large:

$$z_{S\cup\{a\}} \geq \sqrt{\frac{k+1-i}{k+2}} \cdot z_{S_{k-i}} + c''\left(\sum_{j=0}^{i} \frac{1}{\sqrt{k-j}} \cdot \sqrt{\frac{k+2-j}{k+2}}\right),$$

where $S_{k-i} = \{2^{2^\ell} \mid 1 \leq \ell \leq k-i-1\} \cup \{0,1\}$. Now let $i = [k/2]$, where $[x]$ denotes the greatest integer less than or equal to x. Hence, we have:

$$\begin{aligned} z_{S\cup\{a\}} &\geq \sqrt{\frac{k+1-i}{k+2}} \cdot z_{S_{k-i}} + c''\left(\sum_{j=0}^{[k/2]} \frac{1}{\sqrt{k-j}} \cdot \sqrt{\frac{k+2-j}{k+2}}\right) \\ &\geq \frac{1}{\sqrt{2}} \cdot z_{S_{[k/2]}} + c''([k/2]+1)\frac{1}{\sqrt{2}\sqrt{k}} \\ \text{(for some real number } c''') \quad &\geq c'''\sqrt{k}. \end{aligned}$$

This proves the theorem. □

7. Conclusion

We settle the bound on the expected number of real zeros of a random k-sparse polynomials when the coefficients are independent standard normal random variables. We first showed an $O(\sqrt{k})$ upper bound for an arbitrary set of size k, and then gave an example of set where this bound is tight. We see this as another step towards understanding the number of real zeros of sparse polynomials and related generalizations.

In this article, we considered random variables following independent standard normal distributions. It would be interesting to study other distributions on the coefficients, although we expect analysis to become increasingly difficult as the distributions become more complex.

We also mentioned how the real τ-conjecture is connected to the problem we study and its importance in algebraic complexity. Towards resolving the conjecture, consider the simple setting where f and g are both k-sparse polynomials and we wish to study the number of real zeros of $fg+1$. This is essentially the first case which is non-trivial, unfortunately very little is known and prior techniques seem to fail so far.

Also, there is a vast number of restricted arithmetic circuit models. We invite the community, especially experts on these models, to consider the number of real zeros of univariate polynomials under such restrictions and explore their connections with complexity theoretic lower bounds. It is conceivable that one can find a restriction for which the behavior of the expected number of real zeros is easier to understand than the sparse case and which may lead to new insights towards resolving the aforementioned generalizations, such as the ones considered in the real τ-conjecture.

Acknowledgements

We thank our advisor Markus Bläser for his constant support throughout the work. We thank Vladimir Lysikov for many insightful discussions on the topic. AP thanks Sébastien Tavenas for hosting him at Université Savoie Mont Blanc, Chambéry and for encouraging discussions there.

References

1. Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998, With a foreword by Richard M. Karp. MR 1479636
2. Irénée Briquel and Peter Bürgisser, *The real tau-conjecture is true on average*, CoRR **abs/1806.00417** (2018).
3. Peter Bürgisser, *On defining integers and proving arithmetic circuit lower bounds*, Comput. Complexity **18** (2009), no. 1, 81–103. MR 2505194
4. Peter Bürgisser, Alperen A. Ergür, and Josué Tonelli-Cueto, *On the number of real zeros of random fewnomials*, 2018.
5. Felipe Cucker, Pascal Koiran, and Steve Smale, *A polynomial time algorithm for diophantine equations in one variable*, J. Symb. Comput. **27** (1999), no. 1, 21–29.
6. Jacopo D'Aurizio, *A generalization of the shafer-fink inequality*, 2013.
7. René Descartes, *La géométrie*, Hermann, 1886.
8. Alan Edelman and Eric Kostlan, *How many zeros of a random polynomial are real?*, Bull. Amer. Math. Soc. (N.S.) **32** (1995), no. 1, 1–37. MR 1290398
9. Arno Eigenwillig, *Real root isolation for exact and approximate polynomials using descartes' rule of signs*, Ph.D. thesis, Saarland University, 2008.
10. Tamás Erdélyi, *Extensions of the bloch–pólya theorem on the number of real zeros of polynomials*, Journal de théorie des nombres de Bordeaux **20** (2008), no. 2, 281–287 (en). MR 2477504
11. Paul Erdös and A. C. Offord, *On the number of real roots of a random algebraic equation*, Proc. London Math. Soc. (3) **6** (1956), 139–160. MR 0073870
12. Pavel Hrubes, *On the real τ-conjecture and the distribution of complex roots*, Theory of Computing **9** (2013), 403–411.
13. I. A. Ibragimov and N. B. Maslova, *On the average of real zeroes of random polynomials. i. the coefficients with zero means*, Teor. Veroyatnost. i Primenen. **16** (1971), 229–248.
14. M. Kac, *On the average number of real roots of a random algebraic equation*, Bull. Amer. Math. Soc. **49** (1943), 314–320. MR 7812
15. A. G. Khovanskiĭ, *Fewnomials*, Translations of Mathematical Monographs, vol. 88, American Mathematical Society, Providence, RI, 1991, Translated from the Russian by Smilka Zdravkovska. MR 1108621
16. Pascal Koiran, *Shallow circuits with high-powered inputs*, Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings, 2011, pp. 309–320.
17. Pascal Koiran, Natacha Portier, and Sébastien Tavenas, *A wronskian approach to the real τ-conjecture*, J. Symb. Comput. **68** (2015), 195–214.
18. Pascal Koiran, Natacha Portier, Sébastien Tavenas, and Stéphan Thomassé, *A τ-conjecture for newton polygons*, Foundations of Computational Mathematics **15** (2015), no. 1, 185–197.

19. Hendrik W. Lenstra (Jr.), *Finding small degree factors of lacunary polynomials*, Number Theory in Progress **1** (1999), 267–276.
20. J. E. Littlewood and A. C. Offord, *On the Number of Real Roots of a Random Algebraic Equation*, J. London Math. Soc. **13** (1938), no. 4, 288–295. MR 1574980
21. J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation (iii)*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 277–286.
22. Gregorio Malajovich and J. Maurice Rojas, *High probability analysis of the condition number of sparse polynomial systems*, Theoret. Comput. Sci. **315** (2004), no. 2-3, 524–555. MR 2073064
23. J. Maurice Rojas, *On the average number of real roots of certain random sparse polynomial systems*, The mathematics of numerical analysis (Park City, UT, 1995), Lectures in Appl. Math., vol. 32, Amer. Math. Soc., Providence, RI, 1996, pp. 689–699. MR 1421361
24. Fabrice Rouillier and Paul Zimmermann, *Efficient isolation of polynomial's real roots*, Journal of Computational and Applied Mathematics **162** (2004), no. 1, 33 – 50, Proceedings of the International Conference on Linear Algebra and Arithmetic 2001.
25. Michael Sagraloff, *A near-optimal algorithm for computing real roots of sparse polynomials*, International Symposium on Symbolic and Algebraic Computation, ISSAC '14, Kobe, Japan, July 23-25, 2014, 2014, pp. 359–366.
26. Michael Sagraloff and Kurt Mehlhorn, *Computing real roots of real polynomials*, J. Symb. Comput. **73** (2016), 46–86.
27. Ramprasad Saptharishi, *A survey of lower bounds in arithmetic circuit complexity*, Github survey (2015).
28. Amir Shpilka and Amir Yehudayoff, *Arithmetic circuits: A survey of recent results and open questions*, Foundations and Trends in Theoretical Computer Science **5** (2010), no. 3-4, 207–388.
29. Michael Shub and Steve Smale, *On the intractability of Hilbert's Nullstellensatz and an algebraic version of "*NP ≠ P*?"*, Duke Math. J. **81** (1995), 47–54 (1996), A celebration of John F. Nash, Jr. MR 1381969
30. D. C. Stevens, *The average number of real zeros of a random polynomial*, Communications on Pure and Applied Mathematics **22** (1969), no. 4, 457–477.

8. Appendix

8.1. Recovering the classics: $O(\log n)$ bound in the dense case

In this section, we give a simple proof of Theorem 1.5 using the tools and notations developed in Section 3. To this end, first we prove the following lemma.

Lemma 8.1. *Let S_1 and $S_2 = \{a\}$ be such that $a > \max(S_1)$. Then* $\frac{g_{S_1}}{g_{S_2}}$ has no critical points in $(0,1)$.

Proof. Critical points of $\frac{g_{S_1}}{g_{S_2}}$ are exactly the zeroes of $g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2}$. We have:

$$\begin{aligned} g_{S_2}g'_{S_1} - g_{S_1}g'_{S_2} &= x^{2a}g'_{S_1} - 2ax^{2a-1}g_{S_1} = x^{2a-1}(xg'_{S_1} - 2ag_{S_1}). \\ &= x^{2a-1}\sum_{e\in S_1}(2e-2a)x^{2e}. \end{aligned}$$

which is clearly always negative in $(0,1)$. Thus $\frac{g_{S_1}}{g_{S_2}}$ has no critical points in (0,1). ◻

Theorem 8.2. *If $S = \{0,1,2,\ldots,n\}$ then $z_S \leq \frac{3}{4}\log_2(n)$.*

Proof. We prove it by induction on n. The base case of $n = 1$ is trivially true.

Suppose n is odd i.e. $n = 2a+1$ for some $a \in \mathbb{Z}_+$. In this case S is the independent sum of $\{0,1,\ldots,a\}$ and $\{0,a+1\}$. Therefore by using Lemma 3.8, we know that $z_S \leq z_{\{0,1,\ldots,a\}} + z_{\{0,a+1\}}$. We know, $z_{\{0,a+1\}} = \frac{1}{4}$. By using the induction hypothesis, we know that $z_{\{0,1,\ldots,a\}} \leq \frac{3}{4}\log_2(a)$. Hence $z_S \leq \frac{3}{4}\log_2(a) + \frac{1}{4} \leq \frac{3}{4}\log_2(2a+1)$.

Now consider the case when n is even i.e. $n = 2a$ for some $a \in \mathbb{Z}_+$. We have $S = \{0,1,\ldots,2a-1\} \cup \{2a\}$. By using Lemma 8.1 and Corollary 3.13, we get that

$$z_S \leq z_{\{0,1,\ldots,2a-1\}} + \frac{1}{2} \leq z_{\{0,1,\ldots,a-1\}} + \frac{1}{4} + \frac{1}{2} \leq \frac{3}{4}\log_2(a-1) + \frac{3}{4} \leq \frac{3}{4}\log_2(2a).$$

□

Theorem 8.2 shows that $z_{\{0,1,\ldots n\}} \leq \frac{3}{4}\log_2(n)$, which is worse bound than Theorem 1.5. But asymptotically they are similar.

8.2. Proof of the $O(\sqrt{k}\log(k))$ bound [4] [Theorem 1.6]

Before giving the proof we first draw attention to the following folklore lemma about ℓ_1 and ℓ_2 norms which is used in the proof.

Lemma 8.3. *For all $x \in \mathbb{R}^k$, we have the following inequality between ℓ_1 and ℓ_2 norms of x:*

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{k}\,\|x\|_2\,.$$

Proof. The inequality $\|x\|_2 \leq \|x\|_1$ is trivial. For the second inequality, we use Cauchy-Schwartz to get:

$$\|x\|_1 = \sum_{i=1}^{k} |x_i| = \sum_{i=1}^{k} |x_i| \cdot 1 \leq \left(\sum_{i=1}^{k} x_i^2\right)^{1/2} \left(\sum_{i=1}^{k} 1^2\right)^{1/2} = \sqrt{k}\,\|x\|_2\,.$$

□

The following upper bound on z_S was proven in [4] using the Edelman-Kostlan integral.

Theorem 8.4 (Theorem 1.3 in [4])**.** *Let $S \subseteq \mathbb{N}$ be any set as above with $|S| = k$ then we have*

$$z_S \leq \frac{1}{\pi}\sqrt{k} \cdot \log(k).$$

Proof. We use the inequality as in the proof of Theorem 1.4.

$$z_S \leq \frac{1}{\pi}\int_0^1 \frac{\sqrt{(\|v_S(t)\|_2 \cdot \|v_S'(t)\|_2)^2}}{(\|v_S(t)\|_2)^2}\mathrm{d}t = \frac{1}{\pi}\int_0^1 \frac{\|v_S'(t)\|_2}{\|v_S(t)\|_2}\mathrm{d}t$$

(By using Lemma 8.3)

$$\begin{aligned}
&\leq \frac{1}{\pi}\int_0^1 \sqrt{k}\cdot\frac{\|v_S'(t)\|_1}{\|v_S(t)\|_1}\mathrm{d}t \\
&= \frac{1}{\pi}\sqrt{k}\cdot[\log(\|v_S(t)\|_1)]_0^1 = \frac{1}{\pi}\sqrt{k}\cdot(\log(\|v_S(1)\|_1) - \log(\|v_S(0)\|_1)) \\
&= \frac{1}{\pi}\sqrt{k}\log(k).
\end{aligned}$$

□

Department of Computer Science, Aalto University, Finland. Supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 759557) and by Academy of Finland, under grant number 310415

Email address: gorav.jindal@gmail.com

Max-Planck Institute for Computer Science, Saarland Informatics Campus, Saarbrücken, Germany

Email address: apandey@mpi-inf.mpg.de

Max-Planck Institute for Computer Science, Saarland Informatics Campus, Saarbrücken, Germany

Email address: hshukla@mpi-inf.mpg.de

Department of Computer Science, Saarland Informatics Campus, Saarbrücken, Germany

Email address: s9chziso@stud.uni-saarland.de