How Many Zeros of a Random Sparse Polynomial Are Real?

Gorav Jindal
gorav.jindal@gmail.com
Department of Computer Science, Aalto University
Espoo, Finland

Anurag Pandey
apandey@mpi-inf.mpg.de
Max Planck Institut für Informatik,
Saarland Informatics Campus
Saarbrücken, Germany

Himanshu Shukla
hshukla.math04@gmail.com
Max Planck Institut für Informatik,
Saarland Informatics Campus
Saarbrücken, Germany

Charilaos Zisopoulos
zisopoulos@cs.uni-saarland.de
Department of Computer Science, Saarland University,
Saarland Informatics Campus
Saarbrücken, Germany

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.

CCS CONCEPTS

• **Theory of computation → Algebraic complexity theory**;
• **Mathematics of computing → Continuous functions**.

KEYWORDS

Sparse Polynomials, Real Tau conjecture, Random polynomials

ACM Reference Format:
Gorav Jindal, Anurag Pandey, Himanshu Shukla, and Charilaos Zisopoulos. 2020. How Many Zeros of a Random Sparse Polynomial Are Real?. In *International Symposium on Symbolic and Algebraic Computation (ISSAC '20), July 20–23, 2020, Kalamata, Greece.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3373207.3404031

ISSAC '20, July 20–23, 2020, Kalamata, Greece

ACM ISBN 978-1-4503-7100-1/20/07…$15.00
https://doi.org/10.1145/3373207.3404031

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.

1.1 Zeros of Sparse Polynomials

A lot of the polynomials that we encounter in applications are *sparse*, i.e., their degree is considerably larger than their number of monomials. This motivates studying the question for the sparse polynomials. Descartes' famous rule of signs from the 17th century [6] already sheds some light by bounding the number of non-zero real zeros of a k-sparse $f \in \mathbb{R}[x]$ [1] 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.

In this spirit, Kushnirenko initiated a systematic study of the number of real zeros of systems of multivariate sparse polynomial equations. He coined the term "fewnomials" for sparse polynomials and made a series of hypotheses connecting the number of real zeros of a system of multivariate polynomial equations to the complexity of symbolic description of the same system. We refer the readers to a letter [17] which he wrote to Frank Sottile telling about the story of the genesis of this study. Since the formulation of the hypotheses in late 1970s, there has been a lot of work on bounding the number of real zeros of a system of sparse polynomials, most notably [1] and [10]. See [24] and [9] for surveys on the topic.

In the setting of a single univariate polynomial, however, our understanding seems very limited. For instance, if we consider the first non-trivial generalization, i.e. if we consider polynomials of the form $fg+1$, where f and g are both k-sparse, to the best of our knowledge, no bound better than the one given by Descartes' rule of sign is known. 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] throughout this article, polynomials considered are over reals and have degree n with $n >> k$.

1.2 Connections to algebraic complexity theory: Real Tau Conjecture

Koiran [14] provided a strong motivation for computer scientists to consider generalizations like the ones above in 2011, when he considered the number of real zeros of the sum of products of sparse polynomials. He formulated *the real τ-conjecture* 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 $O(mkt)$. Thus the conjecture claims that a univariate polynomial computed by a depth-4 arithmetic circuit (see [21, 22] 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 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 [23] 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 superpolynomial lower bound on the arithmetic circuit size of the permanent [4] and also implies $\mathrm{P}_{\mathbb{C}} \neq \mathrm{NP}_{\mathbb{C}}$ in the Blum-Shub-Smale model of computation (see [2, 23]). 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 quickly.

While the real τ-conjecture remains open (see [11, 15, 16] for some works towards it), Briquel and Bürgisser [3] 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 $O(mk^2t)$.

1.3 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. 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, where there are no restrictions on the sparsity, thus we have a polynomial f of degree n with all its $n+1$ coefficients as standard normal random variables, has been extensively studied and is well understood. It has been considered among others by Littlewood, Offord, Erdős, Kac, Edelman, Kostlan for various distribution since the 30s (see for instance [7, 8, 12, 18]). For this article, the works in [7, 12] are most relevant, since it was Kac [12] who showed the first $O(\log n)$ bound for the expected number of real zeros for the dense case when the coefficients are standard normal random variables. It seems very surprising that there are so few real zeros in the random case. Edelman and Kostlan [7] gave an alternative, simpler derivation for the same bound, in addition to providing essential insights to the integral and numerous generalizations in a variety of cases.

In the sparse case, there is a line of work considering the case of the multivariate system of random equations (for instance see [13, 19, 20]). 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 [5] which showed that for a random k-sparse univariate polynomial, the expected number of real roots in the standard normal case, is bounded by $\frac{4}{\pi}\sqrt{k}\log k$, where the base of the logarithm is e, as will be everywhere else in this article unless stated otherwise. Thus they show that in this setting, the number of real zeros is much less than the Descartes bound.

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 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. *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. Since our bound in Theorem 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 2. *There exists a sequence of sets $S_k \subset \mathbb{N}$ with $|S_k| = k+2$ such that for $k \geq 3$, $z_{S_k} \geq \frac{\pi-\sqrt{3}}{16\pi}\sqrt{k} + \frac{1}{7}$.*

Theorem 2 shows that the bound obtained in Theorem 1 is tight and cannot be reduced further for an arbitrary, in terms of just the size of S, $S \subset \mathbb{N}$.

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

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

1.4 Proof ideas

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

The formulation allows us to bound $z_{S_1 \uplus S_2}$ in terms of the bounds on z_{S_1} and z_{S_2} (presented in Subsection 2.2). 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 3).

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, that is, the so-called Minkowski sum of sets S_1 and S_2 (presented in Subsection 2.1). Combining the bounds on $z_{S_1+S_2}$ and $z_{S_1 \uplus S_2}$ 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 Minkowski sums of sets (presented in the full version).

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 [5] makes the analysis of the integral simpler.

Finally in Section 5, 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 2.2) on a carefully chosen monomial sequence, thus proving Theorem 2.

1.5 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 4 ([7, 12])**.** *If $S = \{0, 1, 2, \ldots, n\}$ then*

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

Here $C_1 \approx 0.6257358072\ldots$.

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 [5].

Theorem 5 ([5, Theorem 1.3])**.** *Let $S \subseteq \mathbb{N}$ be any set as above with $|\ S\ | = k$ then we have*

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

2 PRELIMINARIES

Since our method builds upon the Edelman-Kostlan method [7] by a novel approach on analyzing their integral, it is essential to look at the method. In order to compute z_S for $S = \{e_1, \ldots, e_k\}$, define a generalization of the moment curve v_S as $v_S(t) := (t^{e_1}, t^{e_2}, \ldots, t^{e_k})$. This allows the following expression for z_S^I:

Theorem 6 ([7], Theorem 3,1)**.** *For all sets $S \subseteq \mathbb{N}$, we have the following equality for z_S^I*

$$z_S^I = \frac{1}{\pi}\int_I \frac{\sqrt{(\|v_S(t)\|_2 \cdot \|v_S'(t)\|_2)^2 - (v_S(t)\cdot v_S'(t))^2}}{(\|v_S(t)\|_2)^2}\mathrm{d}t. \quad (2.1)$$

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

The strength of this method is that the 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 $S = \{0, 1, \ldots, k\}$ and $I = (0, 1)$ and for these values show 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 [5], 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. In order to further improve this bound, we believe it is necessary to analyze the above integral in new ways.

We now give an alternative formulation of the Edelman-Kostlan integral on which our proofs build upon.

Definition 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 2. Let $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be differentiable function such that $g^{-1}(0)$ is finite. Define the function $\mathcal{I}(g) : \mathbb{R} \to \mathbb{R}$,

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

Note that whenever the Edelman-Kostlan integral is well-defined, the conditions on g which make $\mathcal{I}(g)$ well-defined and non-negative are also satisfied. We now give our alternative formulation.

Lemma 1. *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 (2.1) 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 note the following equality for $v_S(t)\cdot v_S'(t)$.

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

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

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

Therefore we can rewrite z_S^I as

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

□

The formulation in Definition 2 yields the following lemma:

Lemma 2. *For two non-negative functions $g_1, g_2 : \mathbb{R} \to \mathbb{R}_{>0}$, we have that $\sqrt{\mathcal{I}(g_1\cdot g_2)} \le \sqrt{\mathcal{I}(g_1)} + \sqrt{\mathcal{I}(g_2)}$.*

Proof. Consider:

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

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

Lemma 2 allows us to give a bound on the integral when $S = S_1 * S_2$, where $*$ corresponds to the operation of either union or Minkowski sum of sets. This bound depends on the integrals associated to the corresponding sets S_1 and S_2.

2.1 Minkowski sum of sets

In this subsection, we upper bound the number of zeroes z_S when S is the Minkowski sum of two collision-free sets $S = A + B$ by the sum of the number of zeroes for the two summands.

Definition 3. For sets $A, B \subseteq \mathbb{N}$, we define the Minkowski 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. *If $S_1, S_2 \subseteq \mathbb{N}$ are two collision-free sets, 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

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

, where the last inequality follows from Lemma 2. □

2.2 Union of sets

In this subsection, we provide an upper bound for another set operation on the support S. Specifically, we want to find upper bounds for $z_{S_1 \uplus S_2}$, here $S_1 \uplus S_2$ denotes the disjoint union of S_1 and S_2. First we state the following proposition which is easy to verify.

Proposition 1. *If $S_1, S_2 \subseteq \mathbb{N}$ are two disjoint sets then $g_{S_1\uplus S_2}(t) = g_{S_1}(t) + g_{S_2}(t)$.*

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

Definition 4. Let $S_1, S_2 \subseteq \mathbb{N}$ be two disjoint sets with $\left(\frac{g_{S_1}}{g_{S_2}}\right)' \ge 0$ at zero. Let $c_1, \ldots, c_m$ (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 $c \in (0, 1)$.

$$\begin{aligned}\gamma_{S_1,S_2}(c) &= \sqrt{\frac{g_{s_1}(c)}{g_{s_2}(c)}}\\ T^i_{S_1,S_2} &:= (-1)^i\left(\arctan(\gamma_{S_1,S_2}(c_{i+1})) - \arctan(\gamma_{S_1,S_2}(c_i))\right)\\ R_{S_1,S_2} &:= \sum_{i=0}^{m} T^i_{S_1,S_2}.\end{aligned}$$

We also state a basic easy to verify technical proposition which will be useful in the proof of the main theorem.

Proposition 2. *The following identity is true for all a, b, c, d:*

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

We may now state the key result of this section.

Lemma 4. *Let $S_1, S_2 \subseteq \mathbb{N}$ be two disjoint sets. Assume that $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is non-negative at zero. Note that at least one of $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ and $\left(\frac{g_{S_2}}{g_{S_1}}\right)'$ is non-negative at zero. Thus, we can always rename accordingly S_1 and S_2 to ensure this is the case. Then we have*

$$z_{S_1\uplus S_2} \le z_{S_1} + z_{S_2} + \frac{1}{\pi}R_{S_1,S_2}.$$

Proof. By using Proposition 1, we know that

$$\begin{aligned}\mathcal{I}(g_{S_1\uplus S_2}) &= \mathcal{I}(g_{S_1}+g_{S_2})\\ &= \frac{g''_{S_1}+g''_{S_2}}{g_{S_1}+g_{S_2}} - \left(\frac{g'_{S_1}+g'_{S_2}}{g_{S_1}+g_{S_2}}\right)^2 + \frac{1}{t}\left(\frac{g'_{S_1}+g'_{S_2}}{g_{S_1}+g_{S_2}}\right)\\ &= \frac{g_{S_1}}{g_{S_1}+g_{S_2}}\cdot\mathcal{I}(g_{S_1}) + \frac{g_{S_2}}{g_{S_1}+g_{S_2}}\cdot\mathcal{I}(g_{S_2})\\ &\quad + \frac{1}{g_{S_1}g_{S_2}}\left(\frac{g_{S_1}{g_{S_2}}' - g_{S_2}{g_{S_1}}'}{g_{S_1}+g_{S_2}}\right)^2\end{aligned}$$

The last equality follows by applying Proposition 2 on $g'_{S_1} = a, g_{S_1} = b, g'_{S_2} = c, g_{S_2} = d$. In order to simplify the notations we denote $\frac{1}{g_{S_1}g_{S_2}}\left(\frac{g_{S_1}g'_{S_2}-g_{S_2}g'_{S_1}}{g_{S_1}+g_{S_2}}\right)^2$ by W^2. Therefore we have

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

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:

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

$$\left(\text{substituting } u := \sqrt{\frac{g_{S_1}}{g_{S_2}}}.\ \text{Here } \gamma = \sqrt{\frac{g_{S_1}(\alpha)}{g_{S_2}(\alpha)}} \text{ and } \eta = \sqrt{\frac{g_{S_1}(\beta)}{g_{S_2}(\beta)}}\right)$$

Therefore $J_{\alpha,\beta} = 2(\arctan(\eta) - \arctan(\gamma))$. For intervals where $\left(\frac{g_{S_1}}{g_{S_2}}\right)'$ is negative, we obtain the same result by 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 $z_{S_1\uplus S_2}$ follows by using the quantities defined in Definition 4. □

3 PROOF OF THEOREM 1: $O(\sqrt{k})$ BOUND

We begin with considering the cases where either $|S| = 1$ or $|S| = 2$. This will be the base to construct an inductive argument for the general case, using Lemma 4.

Lemma 5. *For any singleton set S, we have $\mathcal{I}(g_S) = 0$.*

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

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

Lemma 6. *For all sets S of size two, $z_S = \frac{1}{4}$.*

Proof. Without loss of generality we can assume that $S = \{0, a\}$. 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 f by adding a monomial of degree higher than the degree of f, we can bound the expected number of real zeros of the resulting polynomial in terms of the bound for the same quantity for f.

Lemma 7. *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*

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

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

$$\left(\frac{g_{\{a\}}}{g_S}\right)' = \frac{1}{g_S^2}\left(2ax^{2a-1}\sum_{e\in S}x^{2e} - x^{2a}\sum_{e\in S}2ex^{2e-1}\right) > 0 \qquad (3.1)$$

Therefore $\frac{g_{\{a\}}}{g_S}$ is always increasing in $(0,1)$. Abusing the notation slightly, let W be such that

$$W^2 = \frac{1}{g_S g_{\{a\}}}\left(\frac{g'_{\{a\}}g_S - g'_S g_{\{a\}}}{g_S + g_{\{a\}}}\right)^2$$

(similar to Lemma 4). Hence, we have

$$\begin{aligned}\sqrt{\mathcal{I}(g_{S\cup\{a\}}(t))} &= \sqrt{\frac{g_S\cdot\mathcal{I}(g_S)}{g_S+g_{\{a\}}} + \frac{g_{\{a\}}\cdot\mathcal{I}(g_{\{a\}})}{g_S+g_{\{a\}}} + W^2}\\ &\le \sqrt{\mathcal{I}(g_S(t))} + 0 + |W|\end{aligned}$$

By substituting into the formula for $z_{S\cup\{a\}}$, we get

$$z_{S\cup\{a\}} = \frac{1}{2\pi}\int_0^1 \sqrt{\mathcal{I}(g_{S\cup\{a\}}(t))}\mathrm{d}t$$
$$\leq \frac{1}{2\pi}\cdot\left(\int_0^1 \sqrt{\mathcal{I}(g_S(t))}\mathrm{d}t \quad + \right.$$
$$\left.\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{g'_{\{a\}}g_S - g'_S g_{\{a\}}}{g_S + g_{\{a\}}}\right)\mathrm{d}t\right)$$
$$= z_S + \frac{1}{2\pi}\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{g'_{\{a\}}g_S - g'_S g_{\{a\}}}{g_S + g_{\{a\}}}\right)\mathrm{d}t$$

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

$$\int_0^1 \frac{1}{\sqrt{g_S g_{\{a\}}}}\left(\frac{g'_{\{a\}}g_S - g'_S g_{\{a\}}}{g_S + g_{\{a\}}}\right)\mathrm{d}t = 2\int_\alpha^\beta \left(\frac{1}{1+u^2}\right)\mathrm{d}u,$$

where $\alpha = \sqrt{\frac{g_{\{a\}}(0)}{g_s(0)}} = 0$ and $\beta = \sqrt{\frac{g_{\{a\}}(1)}{g_s(1)}} = \frac{1}{\sqrt{k}}$. Note that the integrand is the derivative of $\arctan(u)$, we now have

$$2\int_\alpha^\beta \left(\frac{1}{1+u^2}\right)\mathrm{d}u = 2\left(\arctan\left(\frac{1}{\sqrt{k}}\right) - \arctan(0)\right) = 2\arctan\left(\frac{1}{\sqrt{k}}\right)$$

Hence

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

We may now prove a slightly stronger version of Theorem 1.

Theorem 7 (Theorem 1 restated)**.** *Let $S \subseteq \mathbb{N}$ be a set with $0 \in S$ and $|\ S\ | = k$. Then $z_S \leq \frac{1}{4} + \frac{2}{\pi}(\sqrt{k-1}-1) \leq \frac{2}{\pi}\cdot\sqrt{k-1}$.*

Proof. If $k \leq 2$ then the results follows from Lemma 6. So assume $k > 2$. By using Lemmas 6 and 7, we may always add the highest element iteratively and obtain that

$$z_S \leq \frac{1}{4} + \frac{1}{\pi}\sum_{i=2}^{k-1}\arctan\left(\frac{1}{\sqrt{i}}\right)$$

We use the following well-known inequality

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

This implies that

$$z_S - \frac{1}{4} \leq \frac{1}{\pi}\sum_{i=2}^{k-1}\frac{1}{\sqrt{i}} \leq \frac{1}{\pi}\sum_{i=2}^{k-1}\frac{1}{\sqrt{i}}$$
$$\leq \frac{1}{\pi}\int_1^{k-1}\sqrt{\frac{1}{x}}\mathrm{d}x = \frac{2}{\pi}(\sqrt{k-1}-1).$$

Hence the claimed bound follows. $\square$

4 PROOF OF THEOREM 3: ROOTS CONCENTRATE AROUND 1

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

Proposition 3. *For all $t \in (0,1)$, we have*

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

Proof. First use the following well-known equality

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

to 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

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

$$= \sqrt{\frac{1}{(1-t^2)^2} + \frac{4t^2}{(1-t^2)^3}} \leq \frac{1}{1-t^2} + \frac{2t}{(1-t^2)^{\frac{3}{2}}} \qquad \square$$

We now give the proof of Theorem 3.

Proof of Theorem 3. Without loss of generality , we assume that $0 \in S$, therefore $\|v_S(t)\|_2 \geq 1$ for all $t \in \mathbb{R}$. By using the equality in Theorem 6 and also by ignoring the second term in the numerator in Equation (2.1), we get the following inequality for z_S

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

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

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

$$= \frac{1}{\pi}\left(\frac{1}{2}\log\left(\frac{2-\epsilon}{\epsilon}\right) + \frac{2}{\sqrt{\epsilon(2-\epsilon)}} - 2\right)$$
$$\leq \frac{1}{2\pi}\left(\log\left(\frac{2}{\epsilon}\right) + \frac{4}{\sqrt{\epsilon}} - 4\right). \qquad \square$$

5 THE LOWER BOUND

In this section we will come up with a sequence of sets $(S_k)_{k\geq 1}$ such that the expected number of real zeros of the corresponding polynomials is lower bounded by $\Omega(\sqrt{k})$, for large enough k.

Lemma 8. *Suppose $S = \{e_1, e_2, \ldots, e_k\}$ with $e_k = \max(S)$ and $\epsilon > 0$, then $z_S^{(1-\epsilon,1)} \leq \frac{\epsilon\sqrt{k}e_k}{\pi}$.*

Proof. We have:

$$z_S^{(1-\epsilon,1)} \leq \frac{1}{\pi}\int_{1-\epsilon}^{1} \frac{\sqrt{(\|v_S(t)\|_2 \cdot \|v_S'(t)\|_2)^2}}{(\|v_S(t)\|_2)^2}\mathrm{d}t = \frac{1}{\pi}\int_{1-\epsilon}^{1} \frac{\|v_S'(t)\|_2}{\|v_S(t)\|_2}\mathrm{d}t$$
$$\leq \frac{1}{\pi}\int_{1-\epsilon}^{1} \|v_S'(t)\|_2 \,\mathrm{d}t = \frac{1}{\pi}\int_{1-\epsilon}^{1} \left(\sum_{i=1}^{k}\left(e_i^2 t^{2e_i-2}\right)\right)^{\frac{1}{2}} \mathrm{d}t$$
$$\leq \frac{1}{\pi}\int_{1-\epsilon}^{1} \left(ke_k^2\right)^{\frac{1}{2}} \mathrm{d}t = \frac{1}{\pi}\int_{1-\epsilon}^{1} \left(\sqrt{k}e_k\right)\mathrm{d}t = \frac{\epsilon}{\pi}\left(\sqrt{k}e_k\right) \qquad \square$$

Remark 2. Thus, we can have z_S^I arbitrarily small, for a small enough ϵ. This fact will be crucial in the proof of Theorem 2. Further, Lemma 8 can be viewed as a supplementary result to Theorem 3. Theorem 3 implies that most of the roots lie in $(0, 1-\epsilon)$, if ϵ is allowed to be arbitrarily small. Lemma 8 gives a precise formulation of this fact.

5.1 Proof of Theorem 2

From now on we will assume that $S = \{0,1\} \bigcup \{2^{2^i} \mid 1 \leq i \leq k-1\}$ 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 7.

Lemma 9. *Let W be as in the proof of Lemma 4, then we have $\int_{1-\frac{1}{2a}}^{1} |W|\,\mathrm{d}t \geq 2\left(\arctan\left(\frac{1}{4\sqrt{k}}\right)\right)$.*

Proof. Using the computation in the proof of Lemma 4 we have

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

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 well-known inequalities $\left(1-\frac{1}{n}\right)^n \leq \frac{1}{e}$ (for any $n \in \mathbb{N}$) and $(1+x)^r \geq 1+rx$ if $x \geq -1$ and $r > 1$, we have, for large enough k

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

Therefore, $\arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right) \leq \arctan\left(\sqrt{\frac{1}{e}}\sqrt{\frac{1}{k}}\right)$. which gives

$$2\left(\arctan\left(\frac{1}{\sqrt{k+1}}\right) - \arctan\left(\sqrt{\frac{g_{\{a\}}(1-\frac{1}{2a})}{g_S(1-\frac{1}{2a})}}\right)\right)$$
$$\geq 2\arctan\left(\frac{1}{\sqrt{k+1}}\right) - 2\arctan\left(\sqrt{\frac{1}{e}}\sqrt{\frac{1}{k}}\right)$$
$$\geq 2\left(\arctan\left(\frac{\frac{1}{\sqrt{k+1}} - \frac{1}{e\sqrt{k}}}{1+\frac{1}{e\sqrt{k(k+1)}}}\right)\right) \geq 2\left(\arctan\left(\frac{1}{4\sqrt{k}}\right)\right) \qquad \square$$

For proving Theorem 2 we will again resort to our idea of monomial-wise construction of the 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 that

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

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, then lower bound both integrals.

Proof of Theorem 2. Recall from Lemma 5 that $\mathcal{I}(g_{\{a\}}) = 0$. Therefore we have

$$z_{S\cup\{a\}} = \frac{1}{2\pi}\left(\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S) + 0 + W^2}\mathrm{d}t\right)$$
$$\geq \frac{1}{2\pi}\left(\int_0^{1-\frac{1}{2a}} \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\mathrm{d}t + \int_{1-\frac{1}{2a}}^{1} |W|\,\mathrm{d}t\right)$$
$$= \frac{1}{2\pi}\int_0^1 \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\mathrm{d}t$$
$$- \frac{1}{2\pi}\int_{1-\frac{1}{2a}}^{1} \sqrt{\frac{g_S}{g_S+g_{\{a\}}}\cdot\mathcal{I}(g_S)}\mathrm{d}t + \frac{1}{2\pi}\int_{1-\frac{1}{2a}}^{1} |W|\,\mathrm{d}t$$
$$\geq \frac{1}{2\pi}\sqrt{\frac{k+1}{k+2}}\int_0^1 \sqrt{\mathcal{I}(g_S)}\mathrm{d}t - \frac{1}{2\pi}\int_{1-\frac{1}{2a}}^{1} \sqrt{\mathcal{I}(g_S)}\mathrm{d}t$$
$$+ \frac{1}{2\pi}\int_{1-\frac{1}{2a}}^{1} |W|\,\mathrm{d}t \qquad \left(\frac{g_{\{a\}}}{g_S} \text{ is increasing (Equation (3.1))}\right)$$
$$= \sqrt{\frac{k+1}{k+2}}\, z_S + \frac{1}{2\pi}\left(-\int_{1-\frac{1}{2a}}^{1} \sqrt{\mathcal{I}(g_S)}\mathrm{d}t + \int_{1-\frac{1}{2a}}^{1} |W|\mathrm{d}t\right)$$

Now by using Lemma 8 with $\epsilon = \frac{1}{2a}$, Lemma 9 and the inequality $\frac{\pi}{4}x < \arctan(x)$ for $0 < x < 1$, we have

$$
\begin{aligned}
z_{S\cup\{a\}} &\geq \sqrt{\frac{k+1}{k+2}}\, z_S + \frac{1}{2\pi}\int_{1-\frac{1}{2a}}^{1} |W|\,\mathrm{d}t - z_S^{(1-\frac{1}{2a},1)} \\
&\geq \sqrt{\frac{k+1}{k+2}}\, z_S + \frac{1}{\pi}\arctan\left(\frac{1}{4\sqrt{k}}\right) - \frac{\sqrt{k+1}}{2\pi 2^{2^{k-1}}} \\
&\geq \sqrt{\frac{k+1}{k+2}}\, z_S + \frac{1}{\sqrt{k}}\left(\frac{1}{16} - \frac{\sqrt{k}\sqrt{k+1}}{2\pi 2^{2^{k-1}}}\right) \\
&\geq \sqrt{\frac{k+1}{k+2}}\, z_S + \frac{\pi-\sqrt{3}}{16\pi}\frac{1}{\sqrt{k}.} \qquad \text{(assuming } k \geq 3\text{)} \\
&\geq \sqrt{\frac{2}{k+2}}\cdot z_{\{0,1\}} + \frac{\pi-\sqrt{3}}{16\pi}\left(\sum_{j=0}^{k-1}\frac{1}{\sqrt{k-j}}\cdot\sqrt{\frac{k+2-j}{k+2}}\right) \\
&\qquad \text{(iterating k-1 times)} \\
&\geq \sqrt{\frac{2}{k+2}}\cdot z_{\{0,1\}} + \frac{\pi-\sqrt{3}}{16\pi}\left(\sum_{j=0}^{k-1}\frac{1}{\sqrt{k+2-j}}\cdot\sqrt{\frac{k+2-j}{k+2}}\right) \\
&\geq \frac{\pi-\sqrt{3}}{16\pi}\sqrt{k} + \frac{1}{7} \qquad \left(\text{using } z_{\{0,1\}} = \tfrac{1}{4}\right) \quad \square
\end{aligned}
$$

6 CONCLUSION

We settle the bound on the expected number of real zeros of a random k-sparse polynomial 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 the 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 experts 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.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Frédéric Bihan and Frank Sottile. 2007. New fewnomial upper bounds from Gale dual polynomial systems. *Mosc. Math. J.* 7, 3 (2007), 387–407, 573. https://doi.org/10.17323/1609-4514-2007-7-3-387-407

[2] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. 1998. *Complexity and real computation.* Springer-Verlag, New York. xvi+453 pages. https://doi.org/10.1007/978-1-4612-0701-6 With a foreword by Richard M. Karp.

[3] Irénée Briquel and Peter Bürgisser. 2020. The real tau-conjecture is true on average. *Random Structures & Algorithms* (2020). https://doi.org/10.1002/rsa.20926 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20926

[4] Peter Bürgisser. 2009. On defining integers and proving arithmetic circuit lower bounds. *Comput. Complexity* 18, 1 (2009), 81–103. https://doi.org/10.1007/s00037-009-0260-x

[5] Peter Bürgisser, Ergür Alperen A., and Josué Tonelli-Cueto. 2019. On the Number of Real Zeros of Random Fewnomials. *SIAM Journal on Applied Algebra and Geometry* 3, 4 (2019), 721–732.

[6] René Descartes. 1886. *La géométrie.* Hermann.

[7] Alan Edelman and Eric Kostlan. 1995. How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc. (N.S.)* 32, 1 (1995), 1–37. https://doi.org/10.1090/S0273-0979-1995-00571-9

[8] Paul Erdös and A. C. Offord. 1956. On the number of real roots of a random algebraic equation. *Proc. London Math. Soc. (3)* 6 (1956), 139–160. https://doi.org/10.1112/plms/s3-6.1.139

[9] Boulos El Hilany. 2016. *Géométrie Tropicale et Systèmes Polynomiaux.* Ph.D. Dissertation. LAMA, Universite Savoie Mont Blanc et de Universite Grenoble Alpes.

[10] A. G. Hovanskiĭ. 1980. A class of systems of transcendental equations. *Dokl. Akad. Nauk SSSR* 255, 4 (1980), 804–807.

[11] Pavel Hrubes. 2013. On the Real τ-Conjecture and the Distribution of Complex Roots. *Theory of Computing* 9 (2013), 403–411. https://doi.org/10.4086/toc.2013.v009a010

[12] M. Kac. 1943. On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* 49 (1943), 314–320. https://doi.org/10.1090/S0002-9904-1943-07912-8

[13] A. G. Khovanskiĭ. 1991. *Fewnomials.* Translations of Mathematical Monographs, Vol. 88. American Mathematical Society, Providence, RI. viii+139 pages. Translated from the Russian by Smilka Zdravkovska.

[14] Pascal Koiran. 2011. Shallow circuits with high-powered inputs. In *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings.* 309–320. http://conference.iiis.tsinghua.edu.cn/ICS2011/content/papers/5.html

[15] Pascal Koiran, Natacha Portier, and Sébastien Tavenas. 2015. A Wronskian approach to the real τ-conjecture. *J. Symb. Comput.* 68 (2015), 195–214. https://doi.org/10.1016/j.jsc.2014.09.036

[16] Pascal Koiran, Natacha Portier, Sébastien Tavenas, and Stéphan Thomassé. 2015. A τ-Conjecture for Newton Polygons. *Foundations of Computational Mathematics* 15, 1 (2015), 185–197. https://doi.org/10.1007/s10208-014-9216-x

[17] A. Kushnirenko. 26 February 2008. Letter to Frank Sottile. www.math.tamu.edu/~sottile/research/pdf/kushnirenko.pdf.

[18] J. E. Littlewood and A. C. Offord. 1938. On the Number of Real Roots of a Random Algebraic Equation. *J. London Math. Soc.* 13, 4 (1938), 288–295. https://doi.org/10.1112/jlms/s1-13.4.288

[19] Gregorio Malajovich and J. Maurice Rojas. 2004. High probability analysis of the condition number of sparse polynomial systems. *Theoret. Comput. Sci.* 315, 2-3 (2004), 524–555. https://doi.org/10.1016/j.tcs.2004.01.006

[20] J. Maurice Rojas. 1996. On the average number of real roots of certain random sparse polynomial systems. In *The mathematics of numerical analysis (Park City, UT, 1995).* Lectures in Appl. Math., Vol. 32. Amer. Math. Soc., Providence, RI, 689–699.

[21] Ramprasad Saptharishi. 2015. A survey of lower bounds in arithmetic circuit complexity. *Github survey* (2015).

[22] Amir Shpilka and Amir Yehudayoff. 2010. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science* 5, 3-4 (2010), 207–388. https://doi.org/10.1561/0400000039

[23] Michael Shub and Steve Smale. 1995. On the intractability of Hilbert's Nullstellensatz and an algebraic version of "NP ≠ P?". *Duke Math. J.* 81 (1995), 47–54 (1996). https://doi.org/10.1215/S0012-7094-95-08105-8 A celebration of John F. Nash, Jr.

[24] Frank Sottile. 2011. *Real solutions to equations from geometry.* University Lecture Series, Vol. 57. American Mathematical Society, Providence, RI. x+200 pages. https://doi.org/10.1090/ulect/057