Lower bounds on the sum of 25^{th}-powers of univariates lead to complete derandomization of PIT

Pranjal Dutta *　　Nitin Saxena †　　Thomas Thierauf ‡

Abstract

We consider the univariate polynomial $f_d := (x+1)^d$ when represented as a sum of *constant*-powers of univariate polynomials. We define a natural measure for the model, the *support-union*, and conjecture that it is $\Omega(d)$ for f_d.

We show a stunning connection of the conjecture to the two main problems in algebraic complexity: Polynomial Identity Testing (PIT) and VP vs. VNP. Our conjecture on f_d implies blackbox-PIT in P. Assuming the Generalized Riemann Hypothesis (GRH), it also implies VP $\neq$ VNP. No such connection to PIT, from lower bounds on constant-powers representation of polynomials was known before. We establish that studying the expression of $(x+1)^d$, as the sum of 25^{th}-powers of univariates, suffices to solve the two major open questions.

In support, we show that our conjecture holds over the *integer ring* of any number field. We also establish a connection with the well-studied notion of matrix *rigidity*.

2012 ACM CCS concept: Theory of computation - Algebraic complexity theory, Problems, reductions and completeness, Pseudorandomness and derandomization; Computing methodologies - Algebraic algorithms; Mathematics of computing - Combinatoric problems.
Keywords: hitting set, circuit, univariate polynomial, powers, squares, VP vs VNP, PIT, matrix rigidity, lower bound, monomials, support, CH, #P/Poly.

1 Introduction

Algebraic circuits provide a way to study computation. Here, the complexity classes contain multivariate polynomial families instead of languages. An *algebraic circuit* is a natural model to represent a polynomial compactly; for definition see Section 2.

The class VP contains the families of n-variate polynomials of degree $\mathsf{poly}(n)$ over $\mathbb{F}$, computed by circuits of $\mathsf{poly}(n)$-size. The class VNP can be seen as a non-deterministic analog of the class VP. Informally, it contains the families of n-variate polynomials that can be written as an exponential sum of polynomials in VP; for formal definitions, see Section 2. VP is contained in VNP and it is believed that this containment is strict (Valiant's Hypothesis [Val79a]). For more details see, [Mah14, SY10, BCS13]. Unless specified otherwise, we consider field $\mathbb{F} = \mathbb{Q}$ (or finite field with 'large' characteristic).

The interplay between proving lower bounds and derandomization is one of the central themes in complexity theory [NW94]. In algebraic complexity theory, the derandomization question asks for an efficient deterministic algorithm for *Polynomial Identity Testing* (PIT), i.e. to

*Chennai Mathematical Institute, India (& CSE, IIT Kanpur), pranjal@cmi.ac.in
†CSE, Indian Institute of Technology, Kanpur, nitin@cse.iitk.ac.in
‡Aalen University, Germany, thomas.thierauf@uni-ulm.de

test whether a given algebraic circuit computes the *identically* zero polynomial [KI03]. *Blackbox*-PIT asks for an algorithm to test the zeroness of a given algebraic circuit via mere *query access*. Finding a deterministic polynomial time algorithm for PIT for either version is a long-standing open question.

Since a circuit of size s can have $\exp(s)$ many monomials, we cannot hope to solve PIT in polynomial time by computing the polynomial explicitly. But since evaluation of the polynomial at a point is efficient and a non-zero polynomial evaluated at a random point is non-zero with high probability (by the *Polynomial Identity Lemma* [Ore22, DL78, Zip79, Sch80]), one gets a randomized polynomial time algorithm for PIT. For more details on PIT, see the surveys [Sax09, Sax14, SY10, KS19] or review articles [Wig17, Mul12]. The problem also naturally appears in the algebraic-geometry approaches to the $\mathsf{P} \neq \mathsf{NP}$ question, e.g. [Mul17, Muk16, GMQ16, Gro15, Mul12].

One important direction, from hardness to derandomization, is to design deterministic PIT algorithms for small circuits assuming access to *explicit hard polynomials*. Most of the constructions use the concept of *hitting-set generator* (HSG), which usually incorporates the notion of *combinatorial designs*; these are large uniform set families with small pairwise intersections. Very recent work discovered that PIT is amenable to the phenomenon of *bootstrapping* (of variables) [AGS19, KST19]. Finally, Guo et al. [GKSS19] came up with a HSG without designs and showed that hardness of *constant* (≥ 4) variate polynomials can be used to solve PIT in general.

The classical *Waring problem* asks for a number k whether there exists a number $g(k)$ such that every natural number can be written as the sum of $g(k)$-many k-th powers of numbers. Some celebrated examples are $g(2) = 4$ [Dix64] and $g(3) = 9$ [Kem12]. Later, many variants of Waring's problem for polynomials have been studied using real/complex analytic tools [FOS12, CCG12, BT15]. The *sum-of-squares* problem (SOS) is to represent polynomials as sum of squares. It has many applications in optimization and control theory, see [Lau09, BM16]. Roughly speaking, we want to relate variants of SOS to PIT or to lower bounds. Towards that, we create a new framework and take the first step. Theorems 1 & 3 below state:

> *If $(x+1)^d$ written as sum of $o(d)$ many 25^{th}-powers of univariates requires $\Omega(d)$ many distinct monomials, then blackbox-PIT $\in \mathsf{P}$, and, assuming GRH, we have $\mathsf{VP} \neq \mathsf{VNP}$.*

Prior lower bounds for univariate polynomials. It is known that the computation of most of the polynomials of degree d requires $\Omega(d)$ many arithmetic operations [Mot55, Bel58]. For explicit polynomials, $\sum_{i=0}^{d} \sqrt{p_i}\, x^i$ requires circuits of size $\Omega\left(\sqrt{d/\log d}\right)$, where p_i is the i-th prime number [BCS13, Cor.9.4]. For integral coefficients, the polynomial $\sum_{i=0}^{d} 2^{2^i} x^i$ requires circuits of size $\Omega\left(\sqrt{d/\log d}\right)$ [Str74].

Such polynomials can be converted to *exponentially hard* multilinear polynomial $f_n(\boldsymbol{x})$. Unfortunately, such seemingly *strong* lower bounds are insufficient to separate VP and VNP; because the polynomial families turn out to be *non-explicit*, in particular, f_n may not be in VNP. Thus the hardness alone does not resolve VP vs VNP (see [HS80, Bür13]).

The *Pochhammer-Wilkinson* polynomial, $P_d(x) := \prod_{i=1}^{d}(x-i)$, is conjectured to be hard, i.e. $\mathrm{size}(P_d) \geq \Omega(d)$. Such hardness would imply $\mathsf{VP} \neq \mathsf{VNP}$, assuming GRH [Bür09, Cor.4.2]. This is also related to the famous τ-conjecture [SS+95] about integral roots and its real variants in algebraic complexity [Koi11, KPTT15].

Another way to separate VP and VNP is to show lower bounds of the top-fan-in of an explicit polynomial when written as *sum of powers*. In particular, Koiran [Koi11] implicitly showed that if there exists a univariate polynomial $f_d(x)$ of degree d such that any representation of the form $f_d(x) = \sum_{i=1}^{s} Q_i^{e_i}$, where $\mathrm{sparsity}(Q_i) \leq t$ and arbitrary e_i's, requires $s \geq (d/t)^{\Omega(1)}$, then $\mathsf{VP} \neq \mathsf{VNP}$. The proof applies the depth-4 reduction [AV08, Koi12, GKKS13, Tav15] to flatten a circuit. In the case of $\deg(Q_i) \le t$, a lower bound of $s \ge \Omega(\sqrt{d/t})$ is indeed known [KKPS15]. For $\deg(Q_i) \le 1$, the bound $s \ge \Omega(d)$ has been established for certain polynomials; using the concept of *Birkhoff Interpolation* [GMK17, KPGM18].

The above lower bound connections do *not* give poly-time blackbox-PIT. However, some of them do give conditional *quasi*-poly-time blackbox-PIT [AV08, Bür09, Koi11, Koi12, Tav15].

1.1 New measure and our conjecture

For a polynomial $f(x) \in R[x]$ over a ring R, and a positive integer r, we say that f is computed as a *sum of r-th powers* if we can write

$$f = \sum_{i=1}^{s} c_i \, \ell_i^r \,, \tag{1}$$

for some s, where $c_i \in R$ and $\ell_i(x) \in R[x]$. Interestingly, for any fixed $r \in \mathbb{N}$, the sum of r-th powers is a *complete model* for $R = \mathbb{F}$, a field of characteristic zero (resp. large), see Lemmas 9 and 22.

A natural complexity measure in (1) is the *support-union size*, namely the number of distinct monomials in the representation, $\left|\bigcup_{i=1}^{s} \operatorname{supp}(\ell_i)\right|$ where *support* $\operatorname{supp}(\ell)$ denotes the set of nonzero monomials in the polynomial ℓ. The *support-union size of f* with respect to r and s, denoted $U_R(f, r, s)$ is defined as the minimum support-union size when f is written in the form (1), and ∞, if no such representation exists. Note that s is the top fan-in when (1) is considered as a circuit.

An easy counting argument shows that $U_R(f, r, s) \ge \Omega(|\operatorname{supp}(f)|^{1/r})$, for all s. Note that $|\operatorname{supp}(f)| \le \deg(f) + 1$. We consider the polynomial family $f_d := (x+1)^d$ of degree d. Hence, in this case actually $|\operatorname{supp}(f)| = d + 1$. We want to investigate how close $U_R(f_d, r, s)$ gets to d.

- For $s = 1$, if $r \mid d$, then we have $U_{\mathbb{F}}(f_d, r, 1) \le d/r + 1$, because $(x+1)^d = (x+1)^{(d/r)\cdot r}$.
- For $s = 2$, we show that $U_{\mathbb{F}}(f_d, r, 2) \ge d/r + 1$ (Theorem 25).
- (Small s). For $s = r + 1$ and *any* d, we show that $U_{\mathbb{F}}(f_d, r, r+1) \le d/r + r$ (Lemma 21).
- (Large s). For $s \ge c \cdot (d+1)$ for *any* $c > r$, we show that $U_{\mathbb{F}}(f_d, r, s) \le O(d^{1/r})$ (Lemma 22). Thus, for large s, we get $U_{\mathbb{F}}(f_d, r, s) = \Theta(d^{1/r})$, which resolves this case.

We will restrict d to the domain

$$I_r := \{\, r^\ell - 1 \mid \ell \in \mathbb{N} \,\}\,.$$

Let $\mathbb{F}$ be $\mathbb{Q}$, or a finite field of characteristic $> r$. We see an intriguing trade-off between the measure U and the top fan-in s. Motivated from the examples above we conjecture the following.

Conjecture 1 (C1). *There exist positive constants $\delta_1 \le 1$, $\delta_2 \ge 1$ and a constant prime-power r such that $U_{\mathbb{F}}\left(f_d, r, d^{\delta_1}\right) \ge d/r^{\delta_2}$, for all large enough $d \in I_r$.*

Remarks.

1. For $\delta_1 \in (0, 1]$, $s = d^{\delta_1} \le d$. Then the above example for large s does not apply. On the other hand, by picking a large δ_2, the lower bound on U required, is much smaller than d/r.
2. We believe the conjecture to hold for any *large* $d \in \mathbb{N}$ (i.e. beyond I_r). We believe the conjecture to be true for *most* polynomial families, e.g. $f_d := \sum_{i=0}^{d} 2^{i^2} x^i$ or $f_d := \prod_{i=1}^{d}(x - i)$.

3. For the results of this paper, we could even restrict the degrees of ℓ_i, to be $O(d)$, in the sum of r-th powers representation. This might help in proving the conjecture. For details, see Remark 2 at the end of Section 3.1.

4. One can ask for the number of distinct monomials required to *approximate* $f_d(x)$ as a sum of r-th powers. We believe the above conjecture to hold in the approximative computation model as well. See Conjecture C2 and its consequences in Section B.2.

5. We also study a different measure by taking the sparsity-sum (of ℓ_i's); see Conjecture C3.

1.2 Our results

The central theme of this paper is to show interrelations between the conjecture and derandomization/hardness questions in algebraic complexity. Hardness results have often given *efficient* derandomization [AGS19, GKSS19]. Can the *suspected* hardness of $(x+1)^d$ lead to derandomization? Can studying representations like $(x+1)^d = \sum_i \ell_i^{25}$ give efficient PIT? Older results give no inkling of an answer as they needed the powers to be a *growing* function instead of an absolute constant. We demonstrate a positive answer:

Theorem 1 (Conditional PIT)**.** *If Conjecture C1 holds for some $r \geq 25$, then blackbox-PIT $\in$* P.

Remarks. 1. Older *hardness to derandomization* results are mostly based on depth-4 reduction [AV08, Koi12, Tav15], requiring *arbitrarily small but growing $r = \omega(1)$*. This is the first time that constant r model is connected to derandomization.

2. Older results lead to various conditional derandomizations. E.g. *multi*-variate hard polynomials lead to blackbox-PIT $\in$ QP (*quasipoly-time*) [KI03, AGS19]. Recently, Guo et al. [GKSS19] showed that the hardness of a constant k-variate polynomial yields blackbox-PIT $\in$ P, where $k \geq 4$ (see Theorem 10). Now, we improve it to $k = 1$ and show that the hardness of a simple univariate polynomial, in a much weaker model, also translates to complete derandomization.

3. Our choice of $f_d = (x+1)^d$ is mostly because it is simple. Note that one can compute f_d by repeated squaring which yields circuits of size $O(\log d)$. One could also work with more intricate polynomials, e.g. $\prod_{i=1}^{d}(x-i)$ or $\sum_{i=0}^{d} 2^{i^2} x^i$, whose circuit complexity is not clear, but may well be $\Omega(d)$. Showing Conjecture C1 for any of these polynomials would similarly lead us to the parameters in Theorem 1.

4. One can show that the *approximate* version of the conjecture (see Conjecture C2) implies a poly-time hitting-set for $\overline{\mathsf{VP}}$-circuits (Theorem 29).

We do not know whether Conjecture C1 is true over $\mathbb{F} = \mathbb{Q}$. But we show a *strong* lower bound over localized integer rings (e.g. $\mathbb{Z}$) giving substantial evidence for Conjecture C1. For the algebraic number theory terms, see [Lan13]. For any number field K, let $\mathcal{O}_K$ be the *ring of integers* in K, e.g. $\mathbb{Z}$ in $\mathbb{Q}$. Let $\mathbb{P}$ be a prime ideal of $\mathcal{O}_K$, e.g. $\langle p \rangle$ of $\mathbb{Z}$. Define the *localization* $(\mathcal{O}_K)_{\mathbb{P}} := \{r/s \mid r, s \in \mathcal{O}_K, s \notin \mathbb{P}\}$ which is a domain larger than $\mathcal{O}_K$, e.g. $\mathbb{Z}_{\langle p \rangle}$; it has all fractions except the ones like $1/p$. We show that Conjecture C1 is true over $R := (\mathcal{O}_K)_{\mathbb{P}}$, whenever $\mathbb{P} \mid \langle r \rangle_{\mathcal{O}_K}$ (equivalently $\mathbb{P} \supseteq \langle r \rangle_{\mathcal{O}_K}$).

Theorem 2 (Unconditional lower bound)**.** *Fix a prime-power r, any $s \geq 1$, and $f_d(x) := (x+1)^d$. Fix a number field K and its prime ideal $\mathbb{P}$ such that $\mathbb{P} \mid \langle r \rangle_{\mathcal{O}_K}$. Then, $U_{(\mathcal{O}_K)_{\mathbb{P}}}(f_d, r, s) > d$, $\forall d \in I_r$.*

Remark. The lower bound of $d+1$ is *stronger* than d/r^{δ_2} that Conjecture C1 requires. This suggests that constants like $1/r \in \mathbb{Q} = \mathbb{F}$ may help a bit in writing as sum-of-r-th-powers.

We use the *hardness* of $f_d(x)$ to *explicitly* show separation between VP and VNP, assuming GRH (generalized Riemann hypothesis).

Theorem 3 (Conditional l.b.)**.** *If GRH and Conjecture C1 for some $r \geq 25$, hold then* VP $\neq$ VNP*.*

Remarks.

1. It is interesting to note that if Conjecture C1 holds for more intricate polynomial families, e.g. $\sum_{i=0}^{d} 2^{i^2} x^i$, then we get VP $\neq$ VNP without GRH! This has to do with the explicitness of the polynomial family. For details, see Remark 1 at the end of Section 3.3.

2. It is not clear whether $r = 2$ (i.e. sum of squares hardness) gives efficient derandomization, or strong algebraic lower bounds, from our proof technique. However, in the *non-commutative* setting, it is known that strong lower bound on sum-of-squares implies that Permanent is hard [HWY11]. Our framework can be seen as its analog, in the more natural commutative setting.

Connecting the conjecture to matrix rigidity. We restrict ourselves to $r = 2$ and look at the measure $U_{\mathbb{F}}(\cdot)$. We establish an interesting connection to *matrix rigidity*, a well studied pseudo-random property of a matrix. A matrix $A \in \mathbb{F}^{n \times n}$ is (r, s) *rigid* if A cannot be written as a sum $A = R + S$, where R is a matrix of rank r and S is a matrix with at most s non-zero entries. Valiant [Val77] famously proved that if A is computed by a *linear circuit* with bounded fan-in of depth $O(\log n)$ and size $O(n)$, then A is not $(\epsilon \cdot n, n^{1+\delta})$ rigid for every $\epsilon, \delta > 0$; for a simple proof see [SY10, Thm.3.22]. Thus, rigidity could be a way to prove *super-linear* circuit lower bounds; see [AC19, DGW19, Lok09] and the references therein. We show that a lower bound on $U_{\mathbb{F}}(f_d, 2, d)$ is already of great interest.

Theorem 4 (To rigidity)**.** *If Conjecture C1 is true for $r = 2$ and $\delta_1 = 1$ and some $\delta_2 \geq 1$, then, there exists $\delta > 0$ and infinitely many $n \times n$ matrices A_n s.t. A_n is $(n/2^{\delta_2+3}, n^{1+\delta})$ rigid, for any $\delta < 1$.*

We discuss connections to other models and measures in Section 3.5.

1.3 Proof ideas

Proof idea of Theorem 1. The basic idea is to construct a k (=constant) variate polynomial from $f_d := (x+1)^d$, and show that it is hard, assuming Conjecture C1. With appropriate parameters, this hardness will lead us to efficient hitting-set for VP using the recent result of Guo et al. [GKSS19], see Theorem 10. The choice of many constants in the proof is quite subtle. We found it quite surprising that everything goes through with $r = 25$. We do not know how to improve it to a smaller r (unless [VSBR83] improves).

We construct a k-variate polynomial $P_n(\boldsymbol{x})$ of individual degree at most n from f_d, where k depends on r, δ_1, δ_2. The construction is an inverse Kronecker substitution, i.e., we have

$$P_n(x_1, \ldots, x_k) \mapsto P_n(x^{(n+1)^0}, \ldots, x^{(n+1)^{k-1}}) = f_d(x),$$

where d is the *unique* element in $I_r \cap [((n+1)^k - 1)/(r+1), (n+1)^k - 1]$. The important property of this map is that it is a bijection between $\text{supp}(P_n)$ and $\text{supp}(f_d)$.

We prove that $\text{size}(P_n) > d^{1/\mu}$, where $\mu \geq 1$ is a constant which depends on r, δ_1, δ_2. For the sake of contradiction, assume that this is not the case. Then there is a normal-form circuit (see Section 2 for definitions) that computes P_n with only a polynomial blow-up in size. We cut this circuit at the t-th top multiplication layer, where $5^t \leq r < 5^{t+1}$, and compute the top and bottom part as $\Sigma\Pi$-circuits. Thus, we have P_n computed by a circuit of depth 4 with the top multiplicative fan-in 5^t. One can thus write $P_n = \Sigma_i c_i g_i^r$ and show that, with appropriate parameter setting, there are at most d^{δ_1} summands and the support-union $|\cup_i \text{supp}(g_i)| < d/r^{\delta_2}$. As Kronecker substitution does not increase the summand fan-in and support, f_d has a sum-of-r-th-powers representation with 'small' support-union. This contradicts Conjecture C1.

Note that we require $r \geq 25$ because our calculation needs $t \geq 2$. For $t = 1$ our argument would not work: we get the support-union size $\binom{k+kn/2}{k} > (n+1)^k > d$ instead of d/r^{δ_2} (for large enough n and constant k), which does not yield a contradiction.

The coefficients of P_n are simply $\binom{d}{i}$, which can be computed in poly(d)-time. Hence, P_n is both, explicit and hard! Also, the hardness is $d^{1/\mu} \geq \Omega(n^{k/\mu})$, where $\deg(P_n) = O(n)$. Thus, for $k > 3\mu$, we can invoke Theorem 10 and use P_n to construct a poly-time HSG for VP-circuits.

Proof idea of Theorem 2. Let r be a power of a prime $r_0 \geq 2$. If $d = r^\ell - 1$, for some $\ell \in \mathbb{N}$, one can show that $\binom{d}{i} \equiv \pm 1 \pmod{r_0}$, for every $0 \leq i \leq d$. In particular, $(x+1)^d \bmod r_0$ has $d+1$ many coefficients. On the other hand, as the Frobenius map $\phi : x \mapsto x^{r_0}$ is a GF(r_0)-linear endomorphism, $\ell(x)^r \equiv \ell(x^r) \bmod r_0$, for any univariate integral polynomial ℓ. Note that, Frobenius map does *not* change the support. So, $(x+1)^d \equiv \sum c_i \ell_i^r \bmod r_0$ implies that the support-union of ℓ_i's must have size $\geq d+1$; hence the bound follows.

Essentially the same proof works over $(\mathcal{O}_K)_{\mathbb{P}}$, where prime ideal $\mathbb{P} \mid r\mathcal{O}_K$.

Proof idea of Theorem 3. *Unlike* the proof of Theorem 1, here we construct an n (=non-constant) variate *multilinear* polynomial P_n from $f_d := (x+1)^d$. We show that it is 'hard' assuming Conjecture C1.

$P_n(\boldsymbol{x})$ is such that after Kronecker substitution: $P_n(x_1, \ldots, x_n) \mapsto P_n(x^{2^0}, \ldots, x^{2^{n-1}}) = f_d$, where d is the *unique* element in $I_r \cap [(2^n-1)/(r+1), 2^n-1]$. As expected, the map is a bijection between $\text{supp}(P_n)$ and $\text{supp}(f_d)$.

We prove that P_n requires $d^{1/\mu} = 2^{\Omega(n)}$-size circuit, where μ is a constant which depends on r and δ_1. In spirit, this part is similar to that in the proof of Theorem 1. However, there are many differences in the proof details as the parameters of P_n are 'inverted' (i.e. individual degree vs. number of variables). Interestingly, this part would go through even by a slightly weaker version of Conjecture C1 (e.g. support-union $\geq \Omega(d)$ is not fully used).

Now assume that GRH is true and VP = VNP. Then the counting hierarchy (CH) collapses to P/poly (Theorem 6). It is not hard to show that each bit of $\binom{d}{i}$, in the coefficients of f_d, is computable in CH $\subseteq$ P/poly. Thus, using Valiant's criterion, $\{P_n\}_n \in$ VNP = VP; contradicting the $2^{\Omega(n)}$-hardness of P_n proved above from Conjecture C1. So, we conclude VP $\neq$ VNP.

Proof idea of Theorem 4. If A is not $(\epsilon n, n^{1+\delta})$ rigid, then one can show that A can be written as BC, where 'sparse' matrices B and C can have at most $4\epsilon n^2 + 2n^{1+\delta}$ non-zero entries. Now, the idea is to use f_d to construct matrices A_n that cannot be factored thus.

Define $d := n^2 - 1 \in I_2$, $[x]_n := \begin{bmatrix} 1 & x & \cdots & x^{n-1} \end{bmatrix}$, and similarly $[y]_n$. Define polynomial $g_n(x,y)$ such that after Kronecker substitution: $g_n(x,y) \mapsto g_n(x, x^n) = (x+1)^d = f_d$. Finally, define matrix A_n such that $[y]_n A_n [x]_n^T = g_n(x,y)$.

Suppose $A_n = BC$, with $B \in \mathbb{F}^{n \times t}$, $C \in \mathbb{F}^{t \times n}$ and $t := d/2^{\delta_2+1}$ (which specifies ϵ). Then, $[y]_n B\, C\, [x]_n^T = g_n(x,y)$. We deduce that $f_d = \sum_{i \in [t]} \ell_i(x)\, \tilde{\ell}_i(x^n)$, where $([y]_n B)_i =: \tilde{\ell}_i(y)$ and $(C\, [x]_n^T)_i =: \ell_i(x)$. Note that f_d can easily be written as sum of $2t$ squares.

Assuming Conjecture C1, one can show that union of the supports of $\ell_i, \tilde{\ell}_i$ must be 'large', for $i \in [n]$, which ensures that the number of nonzero entries in B and C is 'large'. Therefore, choosing ϵ and δ carefully, A_n *is rigid* with the stated parameters.

2 Preliminaries

Basic notation. Denote the underlying field as $\mathbb{F}$ and assume that it is $\mathbb{Q}, \mathbb{Q}_p$, or their fixed extensions. Our results hold also for finite fields of large characteristic.

Let $[n] = \{1, \ldots, n\}$. For $i \in \mathbb{N}$ and $b \geq 2$, we denote by $\text{base}_b(i)$ the unique k-tuple $(i_1, \ldots, i_k)$ such that $i = \sum_{j=1}^{k} i_j\, b^{j-1}$. In the special case $b = 2$, we define $\text{bin}(i) = \text{base}_2(i)$.

For estimates on binomial coefficients, we use the following standard bound for $1 \leq k \leq n$,

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k . \tag{2}$$

Complexity classes. We assume that the reader is familiar with the standard complexity classes like P, NP, the *polynomial hierarchy* PH, or the counting class $\#\mathsf{P}$ (see for example [AB09]). The *counting hierarchy* is denoted by CH [Wag86]. The class of poly-size circuits can be expressed by the nonuniform *advice class* P/poly.

Matrix rigidity. A matrix A over $\mathbb{F}$ is (r,s)*-rigid*, if one needs to change $> s$ entries in A to obtain a matrix of rank $\leq r$. That is, one *cannot* decompose A into $A = R + S$, where $\text{rank}(R) \leq r$ and $\text{sp}(S) \leq s$, where $\text{sp}(S)$ is the *sparsity of* S, i.e., the number of nonzero entries in S.

Polynomials. For a multivariate polynomial $p \in \mathbb{F}[\boldsymbol{x}]$, where $\boldsymbol{x} = (x_1, \ldots, x_m)$, for some $m \geq 1$, the *support of* p, denoted by $\text{supp}(p)$, is the set of nonzero monomials in p. The *sparsity* or *support size of* p is $|p|_1 = |\text{supp}(p)|$. By $\text{coef}(p)$ we denote the *coefficient vector* of p (in some fixed order). For polynomials $p_1, \ldots, p_s \in \mathbb{F}[\boldsymbol{x}]$, their *span* is the vector space

$$\text{span}_{\mathbb{F}}(p_1, \ldots, p_s) = \{ \sum_{i=1}^{s} c_i p_i \mid c_i \in \mathbb{F}, \text{ for } i = 1, \ldots, s \}.$$

For an exponent vector $\boldsymbol{e} = (e_1, \ldots, e_k)$, we use $\boldsymbol{x}^{\boldsymbol{e}}$ to denote the monomial $x_1^{e_1} \ldots x_k^{e_k}$.

By $\mathbb{F}[x]^{\leq d}$ we denote the $\mathbb{F}$-vector space of univariate polynomials of degree at most d.

Algebraic circuits. An *algebraic circuit* is a layered directed acyclic graph. The leaf nodes are labeled with the input variables $x_1, \ldots, x_n$ and constants from the underlying field $\mathbb{F}$. All the other nodes are labeled as addition and multiplication gates. The root node outputs the polynomial computed by the circuit. Some of the complexity parameters of a circuit are the *size*, the number of edges and nodes, the *depth*, the number of layers, the *fan-in*, the maximum number of inputs to a node, and the *fan-out*, the maximum number of outputs of a node.

For a polynomial f, the size of the smallest circuit computing f is denoted by $\text{size}(f)$, it is the *algebraic circuit complexity* of f. By $\mathcal{C}(n, D, s)$, we denote the set of circuits C that compute n-variate polynomials of degree D such that $\text{size}(C) \leq s$. The *circuit complexity of a family* $\{P_n\}_n$ is $g(n)$, if $\text{size}(P_n) = \Theta(g(n))$.

The class VP contains the families of n-variate polynomials of degree $\mathsf{poly}(n)$ over $\mathbb{F}$, computed by circuits of $\mathsf{poly}(n)$-size. The class VNP can be seen as a non-deterministic analog of the class VP. A family of n-variate polynomials $\{f_n\}_n$ over $\mathbb{F}$ is in VNP if there exists a family of polynomials $\{g_n\}_n$ in VP such that for every $\boldsymbol{x} = (x_1, \ldots, x_n)$ one can write $f_n(\boldsymbol{x}) = \sum_{w \in \{0,1\}^{t(n)}} g_n(\boldsymbol{x}, w)$, for some polynomial $t(n)$ which is called the *witness size*.

VP and VNP have several closure properties. In particular, they are closed under substitution. That is, for a polynomial $f(\boldsymbol{x}, \boldsymbol{y}) \in \mathsf{VP}$ (or VNP), also $f(\boldsymbol{x}, \boldsymbol{y}_0) \in \mathsf{VP}$ (resp. VNP), for any values $\boldsymbol{y}_0$ from $\mathbb{F}$ assigned to the variables in $\boldsymbol{y}$.

Valiant [Val79b] gace a useful *sufficient* condition for a polynomial family $\{f_n(\boldsymbol{x})\}_n$ to be in VNP.

Theorem 5 (Valiant's criterion, [Val79b]). *A family $\{f_n\}_n$ of polynomials is in VNP if there exists $\phi \in \mathsf{P/poly}$ such that for all $\boldsymbol{x} \in \mathbb{F}^n$,*

$$f_n(\boldsymbol{x}) = \sum_{e \in \{0,1\}^n} \phi(\boldsymbol{e})\, \boldsymbol{x}^{\boldsymbol{e}}\,.$$

Valiant's hypothesis and GRH. Valiant conjectured that $\mathsf{VP} \neq \mathsf{VNP}$. Bürgisser [Bür00, Cor.1.2] showed that if Valiant's hypothesis is false and GRH holds, then the polynomial hierarchy collapses. From this, it is not hard to deduce the following.

Theorem 6. *If GRH is true and $\mathsf{VP} = \mathsf{VNP}$, then $\mathsf{CH} \subseteq \mathsf{P/poly}$.*

Over finite fields, GRH is not needed; GRH is required only for $\mathbb{Q}$.

Normal-form algebraic circuits. In our proofs we need some structural results on algebraic circuits, especially *depth reductions* and *hardness to derandomization* results. For completeness, we state them explicitly.

A *normal-form algebraic circuit* is an algebraic circuit $\mathcal{C}$ with the following properties:

1. $\mathcal{C}$ has *alternating* layers of addition and multiplication gates with the root being addition,
2. below each multiplication layer the associated polynomial degree at least *halves*,
3. the fan-in of each multiplication gate is at most 5 (*multiplicative fan-in*), and
4. $\mathrm{depth}(\mathcal{C}) = O(\log d)$, where d is the degree of the polynomial computed by $\mathcal{C}$.

Any circuit can be computed by a normal-form circuit with only *polynomial* blow up in size.

Theorem 7. [VSBR83, AJMV98] *Suppose $f(\boldsymbol{x}) \in \mathbb{F}[\boldsymbol{x}]$ is a polynomial of degree d which can be computed by a circuit $\mathcal{C}$ of size s. Then there exists a* normal-form *circuit $\mathcal{C}'$ of size $O(s^3 d^6)$ that computes f.*

Every polynomial can be computed by circuit of depth 2, however, with exponential size. Let f be an n-variate polynomial of degree d. It has at most $\binom{n+d}{d}$ monomials. This directly yields a $\Sigma\Pi$-circuit of size $\binom{n+d}{d}$.

Fischer's formula. By a formula due to Fischer [Fis94] one can write any monomial as an exponential sum of powers. It requires char $\mathbb{F} = 0$ or large. Also, it fails over $\mathbb{Z}$.

Lemma 8 ([Fis94]). *Let $\mathbb{F}$ be a field of characteristic 0 or $> m$. Any expression of the form $g = \prod_{i \in [m]} g_i$ can be written as $g = \sum_{j \in [2^m]} c_j h_j^m$, where $c_j \in \mathbb{F}$ and $h_j \in \mathrm{span}_{\mathbb{F}}(g_i \mid i \in [m])$, for $j \in [2^m]$.*

Note that the exponent m of the h_j's in Fischer's formula is determined by the number of factors in the product expression. For our purpose, we need to be more flexible with the exponent. The following lemma shows how to rewrite the sum as powers of r, for any $r \geq m$. Note in the proof that the support-union of h does not change in the new representation.

Lemma 9. *Let $\mathbb{F}$ be a field of characteristic 0 or large. Let $h(x) \in \mathbb{F}[x]$ and $0 \le m \le r$. There exist $c_{m,i} \in \mathbb{F}$ and* distinct *$\lambda_i \in \mathbb{F}$, for $0 \le i \le r$, such that*

$$h(x)^m \;=\; \sum_{i=0}^{r} c_{m,i}\,(h(x)+\lambda_i)^r\,. \tag{3}$$

Proof. Consider the polynomial $(h(x)+t)^r$, where t is a new indeterminate different from x. We have

$$(h(x)+t)^r \;=\; \sum_{i=0}^{r} \binom{r}{i} h(x)^i\, t^{r-i}\,.$$

Choose $r+1$ many distinct λ_i's and put $t = \lambda_i$, for $i = 0,1,\ldots,r$. We get $r+1$ many linear equations which can be represented in matrix form $A\boldsymbol{v} = \boldsymbol{b}$, for matrix $A = \left(\binom{r}{j}\lambda_i^{r-j}\right)_{0\le i,j\le r}$, and vectors $\boldsymbol{v} = \left(h^i\right)_{0\le i\le r}$ and $\boldsymbol{b} = \left((h+\lambda_i)^r\right)_{0\le i\le r}$.

Note that except for the binomial factors, A is a Vandermonde matrix. When computing the determinant, one can pull out the binomial factor $\binom{r}{j}$ from the j-th column, for $j = 0,1,\ldots,r$. Then a Vandermonde matrix remains, and hence

$$\det(A) = \prod_{j=0}^{r}\binom{r}{j} \cdot \prod_{0\le i<j\le r}(\lambda_j - \lambda_i) \neq 0\,.$$

Therefore, A is invertible and we have $\boldsymbol{v} = A^{-1}\boldsymbol{b}$.

Let $\boldsymbol{c}_m$ be the $(m+1)$-th row of A^{-1}. Then we have $h(x)^m \;=\; \boldsymbol{c}_m \cdot \boldsymbol{b}$ which is exactly (3). $\square$

Kronecker map and its inverse. Let $p(x_1,\ldots,x_k)$ be a polynomial, where the variables have individual degree bounded by n. The *Kronecker map* $\phi_{k,n}(p)(x)$ yields a univariate polynomial by replacing variable x_i in p by $x^{(n+1)^{i-1}}$, for all $i \in [k]$.

The map has the property that any polynomial with individual degree at most n gets *uniquely* mapped to a univariate polynomial of degree at most $d \;=\; \sum_{i=1}^{k} n(n+1)^{i-1} \;=\; (n+1)^k - 1$ [Kro82].

Next, we consider the inverse map. Let $q(x)$ be a univariate polynomial of degree d. For $k \ge 1$ let $\boldsymbol{x} = (x_1,\ldots,x_k)$ and $n = \lceil (d+1)^{1/k}\rceil - 1$. The *inverse Kronecker map* $\psi_{k,d}(q)(\boldsymbol{x})$ yields a k-variate polynomial by replacing x^i in q by $\boldsymbol{x}^{\mathrm{base}_{n+1}(i)}$, for all $i \in [k]$.

It is easy to see that $\psi_{k,d}$ maps each x^i to a *distinct* k-variate monomial of individual degree $\le n$, for $0 \le i \le d$. Also, we have $\phi_{k,n} \circ \psi_{k,d}(q) = q$ (thus, $\phi_{k,n} \circ \psi_{k,d} = \mathrm{id}$ over $\mathbb{F}[x]^{\le d}$).

Hitting-set generators and deterministic blackbox-PIT from lower bounds. The technical tool to solve blackbox-PIT is to construct an efficient hitting-set generator.

A polynomial map $G : \mathbb{F}^k \longrightarrow \mathbb{F}^n$ given by $G(\boldsymbol{z}) = (g_1(\boldsymbol{z}), g_2(\boldsymbol{z}), \ldots, g_n(\boldsymbol{z}))$ is a *hitting-set generator* (HSG) for a class $\mathcal{C} \subseteq \mathbb{F}[x_1,x_2,\ldots,x_n]$ of polynomials, if for every nonzero $f \in \mathcal{C}$, we have that $f \circ G = f(g_1,g_2,\ldots,g_n)$ is nonzero.

We say that G is *t-time* HSG, if $\mathrm{coef}(g_i)$ can be computed in time t and the maximum degree of g_i is $\le t$.

Given a HSG, one can construct a *hitting-set*, a set H such that a non-zero circuit is non-zero at some points in H. Crucial here is the size of H which depends on the parameters of the HSG. A t-time HSG G gives a $(td)^{O(k)}$ time blackbox-PIT algorithm, for circuits that compute polynomials of degree $\le d$, over popular fields like rationals $\mathbb{Q}$ or their extensions, local fields $\mathbb{Q}_p$ or their extensions, or finite fields $\mathbb{F}_q$. When k is constant, we get a poly-time blackbox-PIT.

Very recently, Guo et al. [GKSS19] showed how to use the hardness of a *constant* variate explicit polynomial family to derandomize PIT. They need the algebraic circuit hardness to be more than d^3; which requires $k \geq 4$ for the family to exist.

Theorem 10. [GKSS19] *Let $P \in \mathbb{F}[\boldsymbol{x}]$ be a k-variate polynomial of degree d such that* $\mathrm{coef}(P)$ *can be computed in* $\mathsf{poly}(d)$*-time. If* $\mathrm{size}(P) > s^{10k+2}\, d^3$*, then there is a* $\mathsf{poly}(s)$*-time* HSG *for* $\mathcal{C}(s,s,s)$*.*

3 Proofs of the main results

In this section, we prove the four main theorems.

3.1 Conjecture C1 to blackbox-PIT: Proof of Theorem 1

Proof of Theorem 1. Let Conjecture C1 be true for some $r \geq 25$, $\delta_1 > 0$, and $\delta_2 \geq 1$. Let k be a *constant* that will be specified later and $\boldsymbol{x} := (x_1, \ldots, x_k)$. For all large $n \in \mathbb{N}$, there exists *exactly* one $d := d(n)$ such that $d \in I_r \cap \left[\left((n+1)^k - 1\right)/(r+1), (n+1)^k - 1\right]$. This follows from the fact that the ratio of two consecutive elements in I_r can be at most $(r^{\ell+1}-1)/(r^{\ell}-1) < r+1$, for $\ell \geq 2$.

Define the polynomial family $P_n(\boldsymbol{x}) := \psi_{k,d}(f_d)$ via the inverse Kronecker map applied to $f_d = (x+1)^d$. From the definition it is clear that P_n is a k-variate polynomial with individual degree at most n, because the individual degree is bounded by $\lceil (d+1)^{1/k} \rceil - 1 \leq n$. Hence, the total degree of P_n is bounded by kn.

Note that $(P_n)_n$ is an *explicit* family of polynomials because its coefficient vector $\mathrm{coef}(P_n)$ can be computed in $\mathsf{poly}(d) = \mathsf{poly}(n)$ time. To see this, observe that for $\boldsymbol{e} = (e_1, \ldots, e_k)$, we have $\mathrm{coef}(\boldsymbol{x}^{\boldsymbol{e}})(P_n) = \binom{d}{e}$, where $e = \sum_{i=1}^{k} e_i (n+1)^{i-1}$. Also, the number of monomials in P_n is $\mathrm{supp}(P_n) = d+1$.

Next we will show the hardness of the polynomial family $(P_n)_n$. Let

$$\mu \;=\; \frac{3}{\frac{\delta_1}{r} - \frac{7}{k}}\,. \tag{4}$$

We want $\mu > 0$. This enforces a condition for k, namely $k > 7r/\delta_1$.

Claim 11 (Hardness of P_n)**.** *C1* $\implies$ $\mathrm{size}(P_n) > d^{1/\mu}$*, for all large enough n.*

Proof of Claim 11. Assume to the contrary that there exists an infinite subset $J \subseteq \mathbb{N}$ such that $\mathrm{size}(P_n) \leq d^{1/\mu}$, for $n \in J$. We will show that Conjecture C1 is false over an infinite subset $J_r = \{\, d(n) \mid n \in J \,\} \subseteq I_r$ which is a contradiction.

Let C be a circuit of size $\leq d^{1/\mu}$ that computes P_n. Thus, by Theorem 7, there exists a normal-form circuit C' of size $s' := d^{3/\mu}\,(kn)^6$. We cut the circuit C' after the t-th layer of multiplication gates from the top, for a constant $t \geq 2$ to be fixed later. This divides C' into two parts, both of them we express as $\Sigma\Pi$-circuits.

- **Top part:** Since the fan-in of each multiplication gate is 5, the top part of the circuit computes a polynomial of degree at most 5^t. The number of variables is bounded by s', the size of the circuit. Hence, the top part can be written as a $\Sigma\Pi$-circuit of size $s_0 := \binom{s'+5^t}{5^t}$.
- **Bottom part:** Since $\deg(P_n) \leq kn$ and the degree at least halves below every multiplication layer, the bottom part computes several k-variate polynomials, each of degree $\leq kn2^{-t}$. So, the bottom part can be written as a $\Sigma\Pi$-circuits of total size $s_1 := \binom{k+kn2^{-t}}{k}$.

When we recombine the $\Sigma\Pi$-circuits of the two parts, we get a $\Sigma^{s_0}\Pi^{5^t}\Sigma\Pi^{kn2^{-t}}$-circuit that computes P_n,

$$P_n = \sum_{i\in[s_0]}\prod_{j\in[5^t]} g_{i,j}\,, \tag{5}$$

where the polynomials $g_{i,j}$ are the ones computed by the bottom part. So $\deg(g_{i,j}) \le kn2^{-t}$. Because the $g_{i,j}$'s have the same k variables as input, their support-union size is bounded by $|\bigcup_{i,j}\operatorname{supp}(g_{i,j})| \le s_1$.

Now we use Fischer's formula (Lemma 8) , to express the product in (5) as a sum of 2^{5^t} powers. Combined with the sum in (5) and renaming the summands, we can write

$$P_n = \sum_{\ell\in[s_0\,2^{5^t}]} c_\ell\, g_\ell^{5^t}\,. \tag{6}$$

where $g_\ell \in \operatorname{span}_{\mathbb{F}}\left(g_{i,j} \mid j\in[5^t]\right)$, for some $i\in[s_0]$, and $c_\ell\in\mathbb{F}$, for $\ell\in\left[s_0\,2^{5^t}\right]$.

Next we use Lemma 9 to adjust the exponent in (6) from 5^t to r. Choose t such that $5^t \le r < 5^{t+1}$. By Lemma 9, there exist $c_{\ell,j},\lambda_j\in\mathbb{F}$ such that $g_\ell^{5^t} = \sum_{j\in[r+1]} c_{\ell,j}\,(g_\ell+\lambda_j)^r$. We plug this into (6) and rename the summands; then we can write

$$P_n = \sum_{i\in[\tilde{s}]} \tilde{c}_i\,\tilde{g}_i^r\,, \tag{7}$$

where $\tilde{s} := s_0\,(r+1)\,2^{5^t}$ and $\tilde{c}_i\in\mathbb{F}$. Note that the polynomials $\tilde{g}_i$ are in the affine space of the above polynomials $g_{i,j}$. Therefore, polynomials $\tilde{g}_i$ are also k-variate and of degree $\deg(\tilde{g}_i)\le kn2^{-t}$, and have the same support-union as the polynomials $g_{i,j}$ in (5). Hence, $|\bigcup_i\operatorname{supp}(\tilde{g}_i)| \le s_1$.

Recall that P_n is defined via the inverses Kronecker map from f_d, i.e., $P_n(\boldsymbol{x}) = \psi_{k,d}(f_d)$. Hence, when we apply the Kronecker map $\phi_{k,n}$ on P_n, we get back f_d,

$$f_d \;=\; \phi_{k,n}(P_n) \;=\; \sum_{i=1}^{\tilde{s}} \tilde{c}_i\,\phi_{k,n}(\tilde{g}_i)^r\,.$$

Since Kronecker substitution maintains the support size, we have $|\bigcup_i\operatorname{supp}(\phi_{k,n}(\tilde{g}_i))| \le s_1$, and therefore $U_{\mathbb{F}}(f_d,r,\tilde{s}) \le s_1$.

We want to show that $\tilde{s} < d^{\delta_1}$ and $s_1 < d/r^{\delta_2}$, for all large enough n. Then we have $U_{\mathbb{F}}(f_d,r,d^{\delta_1}) < d/r^{\delta_2}$, for all large $d\in J_r\subseteq I_r$ which contradicts Conjecture C1.

Bound on s_0. We start by deriving a bound on s_0. By the standard bound on binomial coefficients (2), we have for large enough n

$$\begin{aligned} s_0 \;=\; \binom{s'+5^t}{5^t} \;\le\; \left(e\left(\frac{s'}{5^t}+1\right)\right)^{5^t} \;&<\; \left(3\,\frac{s'}{5^t}\right)^{5^t} \\ &\le\; \left(3\,\frac{d^{\frac{3}{\mu}}\,(kn)^6}{5^t}\right)^{5^t} \\ &\le\; c\,\left(d^{\frac{3}{\mu}}\,n^6\right)^r, \end{aligned} \tag{8}$$

where $c = \left(\frac{3k}{5^t}\right)^{5^t}$ is a constant, and in the last inequality, we used that $5^t \le r$.

Recall that by our choice of d, we have $d \geq \frac{(n+1)^k - 1}{r+1} > \frac{n^k}{r+1}$, and therefore

$$n < (d\,(r+1))^{\frac{1}{k}}. \tag{9}$$

Plugging (9) into (8), we get

$$s_0 \;<\; c\,\left(d^{\frac{3}{\mu}}\,(d\,(r+1))^{\frac{6}{k}}\right)^r \;=\; c'\,d^{\frac{3r}{\mu}+\frac{6r}{k}} \;=\; c'\,d^{\delta_1 - \frac{r}{k}} \tag{10}$$

where $c' = c\,(r+1)^{6r/k}$ is a constant, and in the last equality, we used that $\delta_1 = 3r/\mu + 7r/k$, which follows from (4).

Bound on $\tilde{s}$. With (10), we get the desired estimate for $\tilde{s}$,

$$\tilde{s} \;=\; (r+1)\,2^{5^t}\,s_0 \;<\; (r+1)\,2^{5^t}\,c'\,d^{\delta_1-\frac{r}{k}} \;<\; d^{\delta_1}. \tag{11}$$

For the last inequality note that r, t, c' are constants. Hence we can choose d large enough to fulfill the inequality.

Bound on s_1. Finally, we show that $s_1 < d/r^{\delta_2}$. Again by (2), we have

$$s_1 \;=\; \binom{k + k\,n\,2^{-t}}{k} \;<\; \left(e\,(1 + n\,2^{-t})\right)^k \;\leq\; 3^k\,n^k\,2^{-tk} \;<\; 3^k\,d\,(r+1)\,2^{-tk},$$

Hence, it suffices to show that $3^k\,d\,(r+1)\,2^{-tk} \leq d/r^{\delta_2}$. This is equivalent to $3^k \leq 2^{tk}/(r^{\delta_2}(r+1))$. Because $r < 5^{t+1}$, it suffices to show that

$$3^k \;\leq\; \frac{2^{tk}}{5^{(\delta_2+1)(t+1)}}. \tag{12}$$

Consider the fraction in (12). When we require $k \geq 3(\delta_2+1)$, we have $2^k/5^{(\delta_2+1)} > 1$. Then the fraction is growing with t. Since we assume $t \geq 2$, it then suffices to satisfy (12) for $t = 2$. Then (12) boils down to $125^{\delta_2+1} \leq (4/3)^k$, which is satisfied for $k \geq 17(\delta_2+1)$. Hence, the above calculations holds when we pick $k > \max\,(17(\delta_2+1), 7r/\delta_1)$. This proves Claim 11. $\square$

Form hardness to HSG. We show that by the hardness of P_n from Claim 11, we can fulfill the assumption in Theorem 10 that $\mathrm{size}(P_n) > s^{10k+2}\deg(P_n)^3$, for some appropriate function $s(n)$. Recall that $\deg(P_n) \leq kn$. Define

$$s(n) = n^{\frac{1}{10k+3}}.$$

Then we have

$$s^{10k+2}\deg(P_n)^3 \;\leq\; s^{10k+2}\,(kn)^3 \;=\; n^{\frac{10k+2}{10k+3}}\,(kn)^3 \;=\; k^3\,n^{4-\frac{1}{10k+3}} \;<\; \frac{n^4}{(r+1)^{1/\mu}}. \tag{13}$$

For the last inequality note that k, r, μ are constants. So for large enough n, the inequality will hold.

Recall from (9) that $n^k/(r+1) < d$. Suppose we have the additional property that $4 \leq k/\mu$. Then we can continue (13) by

$$\frac{n^4}{(r+1)^{1/\mu}} \;\leq\; \frac{n^{k/\mu}}{(r+1)^{1/\mu}} \;<\; d^{1/\mu} \;<\; \mathrm{size}(P_n). \tag{14}$$

Equations (13) and (14) give the desired hardness of P_n.

It remains to fulfill the additional requirement $4 \le k/\mu$. We show that that this holds for $k \ge 19r/\delta_1$:

$$\mu \;=\; \frac{3}{\frac{\delta_1}{r} - \frac{7}{k}} \;\le\; \frac{3}{\frac{\delta_1}{r} - \frac{7\delta_1}{19r}} \;=\; \frac{3 \cdot 19r}{11\delta_1} \;<\; \frac{19r}{4\delta_1} \;\le\; \frac{k}{4}.$$

Hence our overall choice for k is $k \ge \max(17(\delta_2+1), 19r/\delta_1)$.

Thus, Theorem 10 gives a $\mathsf{poly}(s)$-time hitting-set generator for $\mathcal{C}(s,s,s)$. Note that s can be any polynomial because one can choose n appropriately and k is independent of n. Hence, blackbox-PIT $\in$ P. $\square$

Remarks. 1. The same proof works for other polynomials like, $\prod_{i\in[d]}(x \pm i)$ or $\sum_{i=0}^{d} 2^{i^2} x^i$. The hardness-proof part does not change at all (assuming the corresponding Conjecture C1). Their explicitness is also clear as their coefficient vector is computable in poly(d)-time. So, the corresponding P_n will be k (=constant) variate and $\mathsf{poly}(n)$-time explicit.

2. Recall the proof notation. As the degree of $\tilde{g}_i$'s is $\le kn2^{-t}$, the degree of $\phi_{k,n}(\tilde{g}_i)$ is $\le (n+1)^{k-1} \cdot kn2^{-t} \;<\; k\cdot(n+1)^k \cdot 2^{-t} \;\le\; k \cdot (d(r+1)+1)\cdot 2^{-t} \;=\; O(d)$ ($\because k, r, t$ are constants). Thus, it suffices to study the representation of f_d as sum-of-r-th powers ℓ_i^r, where $\deg(\ell_i) \le O(d)$; this should lead to the same conclusion as that in Theorem 1.

3. An approximative version, Conjecture C2, leads to an efficient HSG for the class $\overline{\mathsf{VP}}$. The details are discussed in Theorem 29.

3.2 Evidence towards Conjecture C1: Proof of Theorem 2

In the proof of Theorem 2 we consider the support size of f_d modulo a prime r. We prove first that $(x+1)^d \bmod r$ has full support, i.e. $d+1$. We use a celebrated theorem due to Lucas [Luc78].

Theorem 12 (Lucas's Theorem, [Luc78])**.** *For $m, n \in \mathbb{N}$ and a prime p, let*

$$\begin{aligned} m &= m_k p^k + m_{k-1}p^{k-1} + \ldots + m_1 p + m_0 \\ n &= n_k p^k + n_{k-1}p^{k-1} + \ldots + n_1 p + n_0 \end{aligned}$$

be the base-p representation of m and n. Then

$$\binom{m}{n} \;\equiv\; \prod_{i=0}^{k} \binom{m_i}{n_i} \pmod{p}.$$

Lemma 13. *Let $d = r^\ell - 1$, for some prime r and $\ell \in \mathbb{N}$. Then*

$$(x+1)^d \equiv \sum_{k=0}^{d} (-1)^k x^k \pmod{r}. \tag{15}$$

Therefore, we have for the support size $|(x+1)^d \bmod r|_1 = d+1$.

Proof. The base-r representation of d is $d = \sum_{i=0}^{\ell-1}(r-1)\,r^i$. Let $0 \le k \le d$ and write k in base-r representation, $k = \sum_{i=0}^{\ell-1} k_i r^i$.

By Lucas's Theorem, we have

$$\binom{d}{k} \equiv \prod_{i=0}^{\ell-1} \binom{r-1}{k_i} \pmod{r}. \tag{16}$$

Now observe that $\binom{r-1}{k_i} \equiv (-1)^{k_i} \pmod{r}$. This is because $(r-1)(r-2)\cdots(r-k_i) \equiv (-1)^{k_i} k_i!$ $\pmod{r}$, and hence

$$\binom{r-1}{k_i} \equiv \frac{(-1)^{k_i} k_i!}{k_i!} \equiv (-1)^{k_i} \pmod{r}.$$

Plugging this into (16), we get

$$\binom{d}{k} \equiv (-1)^{\sum_{i=0}^{\ell-1} k_i}.$$

Finally observe that $k = \sum_{i=0}^{\ell-1} k_i r^i \equiv \sum_{i=0}^{\ell-1} k_i \pmod{2}$, because r is odd. This proves (15). □

Proof of Theorem 2. Let r be a power of a prime r_0 and $d \in I_r$. Hence, there is an $\ell \in \mathbb{N}$ such that $d = r^\ell - 1$.

By Lemma 13, we have $|(x+1)^d \bmod r_0|_1 = d+1$. Moreover, r_0 does not divide any of the coefficients $\binom{d}{k}$ because $\binom{d}{k} \equiv (-1)^k \pmod{r_0}$, for any $0 \le k \le d$.

Consider the given *prime* ideal $\mathbb{P}$ of $\mathcal{O}_K$ that contains $\langle r \rangle_{\mathcal{O}_K}$, and hence contains $\langle r_0 \rangle_{\mathcal{O}_K}$. Suppose $\binom{d}{j} \in \langle r_0 \rangle_{(\mathcal{O}_K)_\mathbb{P}}$, for some $0 \le j \le d$. Then, simply by ideal definition, there exists $m \in (\mathcal{O}_K)_\mathbb{P}$ such that $\binom{d}{j} = m r_0$. Since r_0 does not divide $\binom{d}{j}$ and $r_0 \in \mathbb{P}$, the quotient $\binom{d}{j}/r_0$ cannot lie in the localization $(\mathcal{O}_K)_\mathbb{P}$, which is a contradiction.

Thus, $\binom{d}{j} \notin \langle r_0 \rangle_{(\mathcal{O}_K)_\mathbb{P}}$, for all $0 \le j \le d$. Whence,

$$\begin{aligned} f_d(x) \;=\; \sum_{i\in[s]} c_i\, \ell_i^r \;\Longrightarrow\; f_d(x) \;&\equiv\; \sum_{i\in[s]} c_i\, \ell_i(x^r) \bmod \langle r_0 \rangle_{(\mathcal{O}_K)_\mathbb{P}} \\ \Longrightarrow\; \Big| \bigcup_{i\in[s]} \mathrm{supp}(\ell_i(x^r)) \Big| \;&\ge\; d+1 \\ \Longrightarrow\; \Big| \bigcup_{i\in[s]} \mathrm{supp}(\ell_i) \Big| \;&\ge\; d+1, \end{aligned}$$

which gives a lower bound on the support-union size as promised. □

Remarks.

1. The fact that $\mathbb{P}$ is a *prime ideal* is crucial in the above proof. This proof works for the polynomial $g := \sum_{i=0}^{d} 2^{i^2} x^i$ as well, as long as r is odd. This is simply because $2^{i^2} \ne 0 \bmod r_0$, for any odd prime r_0. The rest of the proof remains unchanged. For even r, one can work with the alternative $h := \sum_{i=0}^{d} 3^{i^2} x^i$. Conjecture C1 though may still be true for g & h.

2. This also proves that for any prime-power r, for any integer m coprime to r, and for all $d \in I_r$ we have $U_\mathbb{Z}(m f_d, r, \cdot) > d$. This behavior changes when m, r are *not* coprime.

3.3 Conjecture C1 to VP $\neq$ VNP: Proof of Theorem 3

In the proof of Theorem 3 we need the notion of CH-*definable* sequences that we define first.

Let $p(n), q(n)$ be polynomials. Let $a = (a(n,k))_{n\in\mathbb{N},\, k\le 2^{p(n)}}$ be a sequence of nonnegative integers such that $a(n,k)$ has *exponential bitsize*, i.e., $a(n,k) \le 2^{q(n)}$ for all k.

With the sequence, we associate a language that determines the bits of $a(n,k)$ in binary,

$$\mathrm{Bit}(a) = \{\, (1^n, k, j, b) \mid \text{ the } j\text{-th bit of } a(n,k) \text{ equals } b \,\}\,.$$

In case when $k = 1$, we write $a(n)$ as a shorthand for $a(n,1)$.

Definition 14. *The sequence $a = (a(n,k))_{n,k}$ of integers of exponential bitsize is* CH-definable *if* $\mathrm{Bit}(a) \in \mathsf{CH}$.

The sequences of integers that are definable in CH are closed under exponential additions and multiplication [Bür09, Thm.3.10]. Koiran et al. [KP11, Thm.2.14] used the binary version of the same theorem.

Theorem 15. [Bür09, KP11] *Let $p(n)$ be a polynomial and suppose $(a(n,k))_{n\in\mathbb{N},k\le 2^{p(n)}}$ is* CH*-definable. Then the sum- and product-sequences $b(n)$ and $c(n)$ are* CH*-definable, where*

$$b(n) = \sum_{k=0}^{2^{p(n)}} a(n,k) \qquad \text{and} \qquad c(n) = \prod_{k=0}^{2^{p(n)}} a(n,k)\,.$$

We show that the binomial coefficients are definable in CH. The argument is very similar to the proof that the family $\prod_{i\in[d]}(x+i)$ is CH-definable [Bür09, Cor.3.12].

Theorem 16. *Let $p(n)$ be a polynomial and $d_n \le 2^{p(n)}$. The sequence $a(n,k) = \binom{d_n}{k}$ is* CH*-definable.*

Proof. Let $d = d_n$. Consider the identity $(x+1)^d = \sum_{k=0}^{d} \binom{d}{k} x^k$. For $x = 2^d$ we get

$$v(d) \;=\; (2^d+1)^d \;=\; \sum_{k=0}^{d} \binom{d}{k} 2^{dk}\,.$$

Note that $\binom{d}{k} < 2^d$. Thus the bits of $\binom{d}{k}$ in the binary representation of $v(d)$ do not overlap for different k's, Hence the bits of $\binom{d}{k}$ can be read off the bit-vector of $v(d)$. It is therefore sufficient to show that $v(d)$ is definable in CH.

Note that each bit of $2^d + 1$ can be computed in polynomial time. By Theorem 16, we get that $v(d)$ is definable in CH. □

Proof of Theorem 3. Let GRH and Conjecture C1 be true for some $r \ge 25$, $\delta_1 > 0$, and $\delta_2 \ge 1$. For *non-constant* n let $\boldsymbol{x} := (x_1,\dots,x_n)$. For all large $n \in \mathbb{N}$, there exists *exactly* one $d := d(n)$ such that $d \in I_r \cap [(2^n-1)/(r+1), 2^n-1]$. This follows from the fact that the ratio of two consecutive elements in I_r is at most $(r^{\ell+1}-1)/(r^\ell-1) < r+1$, for $\ell \ge 2$. Thus, $n = \Theta(\log d)$.

Define the polynomial family $P_n(\boldsymbol{x}) := \psi_{n,d}(f_d)$ via the inverse Kronecker map applied to $f_d = (x+1)^d$. Note that P_n is an n-variate *multilinear* polynomial. This is ensured because the individual degree $d_n := \lceil (d+1)^{1/n} \rceil - 1 = 1$, because $(2^n-1)/(r+1) \le d \le 2^n - 1$. Hence, P_n has total degree n.

For the sake of contradiction, assume that VP = VNP. First we show that then $(P_n)_n \in \mathsf{VP}$.

Claim 17. $\mathsf{VP} = \mathsf{VNP} \implies (P_n)_n \in \mathsf{VP}$.

Proof of Claim 17. Let $\mathrm{bin}(i) =: (i_1,\dots,i_n)$ so that $i = \sum_{j=1}^{n} i_j\, 2^{j-1}$. By definition,

$$P_n(\boldsymbol{x}) = \sum_{i=0}^{2^n-1} \phi(i)\, \boldsymbol{x}^{\mathrm{bin}(i)}\,,$$

where $\phi(i) := \binom{d}{i}$. Clearly, $\phi(i) < 2^d \leq 2^{2^n-1} < 2^{2^n} - 1$. Write $\phi(i)$ in binary, i.e. $\phi(i) =: \sum_{j=0}^{2^n-1} \gamma_{i,j} 2^j$, where $\gamma_{i,j} \in \{0,1\}$. From Theorem 16, we know that the sequence of coefficients $\phi(i) = \binom{d}{i}$ is CH-definable. This means that the $\gamma_{i,j}$'s are computable in CH, and hence in P/poly, by our assumptions together with Theorem 6.

Introduce new variables $\boldsymbol{y} = (y_1, \ldots, y_n)$ and consider the auxiliary polynomial $\tilde{\phi}_i(\boldsymbol{y}) := \sum_{j=0}^{2^n-1} \gamma_{i,j} \boldsymbol{y}^{\mathrm{bin}(j)}$. Let $\boldsymbol{y}_0 = (2^{2^0}, 2^{2^1}, \ldots, 2^{2^{n-1}})$. Note that $\boldsymbol{y}_0^{\mathrm{bin}(j)} = 2^j$. Therefore $\tilde{\phi}_i(\boldsymbol{y}_0) = \phi(i)$. Now define

$$\tilde{P}_n(\boldsymbol{x}, \boldsymbol{y}) := \sum_{i,j=0}^{2^n-1} \gamma_{i,j} \, \boldsymbol{y}^{\mathrm{bin}(j)} \, \boldsymbol{x}^{\mathrm{bin}(i)} .$$

Then we have $P_n(\boldsymbol{x}) = \tilde{P}_n(\boldsymbol{x}, \boldsymbol{y}_0)$. Since $\gamma_{i,j} \in \mathsf{P/poly}$, we have $(\tilde{P}_n)_n \in \mathsf{VNP} = \mathsf{VP}$, by Valiant's criterion. As VP is closed under substitution we have $(P_n)_n \in \mathsf{VP}$ as well. This proves Claim 17. $\square$

On the other hand, we show next that Conjecture C1 implies that $(P_n)_n \notin \mathsf{VP}$.

Claim 18. *C1* $\implies$ $(P_n)_n \notin \mathsf{VP}$.

Proof of Claim 18. This proof is very similar to the hardness part of Theorem 1, i.e. Claim 11. However, the parameter setting is slightly different (e.g. there is no k here), so we need to go through the details. Let $\mu > 3r/\delta_1$. We prove that $\mathrm{size}(P_n) > d^{1/\mu} = 2^{\Omega(n)}$.

Assume that this is not the case. Then there exists an infinite subset $J \subset \mathbb{N}$ such that $\mathrm{size}(P_n) \leq d^{1/\mu}$, for all $n \in J$. We will show that Conjecture C1 is false over an infinite subset $J_r := \{d(n) \mid n \in J\} \subseteq I_r$ which is a contradiction.

Let C be a circuit of size $\leq d^{1/\mu}$ that computes P_n, for some n. Recall that P_n is multilinear. Hence, by Theorem 7, there exists an equivalent normal-form circuit C' of size $s' := d^{3/\mu} n^6$. Similar as in Claim 11, we cut the circuit C' after the t-th layer of multiplication gates from the top, for a constant t such that $5^t \leq r < 5^{t+1}$. This divides C' into two parts, both of them we express as $\Sigma\Pi$-circuits. Transforming the two parts into $\Sigma\Pi$-circuits, we get a top part of size $s_0 = \binom{s'+5^t}{5^t}$. The bottom part consists of at most s' many circuits of total size $s_1 := \binom{n+n2^{-t}}{n}$.

Then we apply again Fischer's formula and Lemma 9 to write P_n as in (7) (on page 11) as

$$P_n = \sum_{i \in [\tilde{s}]} \tilde{c}_i \, \tilde{g}_i^r ,$$

where $\tilde{s} := s_0 \, (r+1) \, 2^{5^t}$ and $\tilde{c}_i \in \mathbb{F}$, and each $\tilde{g}_i$ is an n-variate polynomial of degree $\leq n/2^t$. The support-union size is $|\bigcup_i \mathrm{supp}(\tilde{g}_i)| \leq s_1$.

Applying the Kronecker map $\phi_{n,1}$ to P_n yields

$$f_d \;=\; \phi_{n,1}(P_n) \;=\; \sum_{i=1}^{\tilde{s}} \tilde{c}_i \, \phi_{n,1}(\tilde{g}_i)^r ,$$

and we have $|\bigcup_i \mathrm{supp}(\phi_{k,n}(\tilde{g}_i))| \leq s_1$, and therefore $U_{\mathbb{F}}(f_d, r, \tilde{s}) \leq s_1$.

We want to show that $\tilde{s} < d^{\delta_1}$ and $s_1 < d/r^{\delta_2}$, for all large enough n. Then we have $U_{\mathbb{F}}(f_d, r, d^{\delta_1}) < d/r^{\delta_2}$, for all large $d \in J_r \subseteq I_r$ which contradicts Conjecture C1.

For a bound on s_0, we have similar to (8) that $s_0 < c \left(d^{\frac{3}{\mu}} n^6 \right)^r$, for large enough n and a constant c different to the one in (8) because there is no k here. Then, with $d = O(\log n)$, we get

$$\tilde{s} \;=\; (r+1) \, 2^{5^t} \, s_0 \;<\; (r+1) \, 2^{5^t} \, c \, d^{3r/\mu} \, (\log d)^{6r} \;<\; d^{\delta_1} .$$

For the last inequality note that r, t, c are constants and $\delta_1 > 3r/\mu$, by our choice of μ.

Finally, we show that $s_1 < d/r^{\delta_2}$, for all large enough d.

$$s_1 \;=\; \binom{n+n2^{-t}}{n2^{-t}} \;<\; \left(e\left(1+2^t\right)\right)^{n2^{-t}} \;<\; \left(\frac{7}{2}\cdot 2^t\right)^{n2^{-t}} . \tag{17}$$

Note that the last expression in (17) decreases with increasing $t \geq 2$. At $t = 2$, it is $14^{n/4} = o(2^n) = o(d)$. For the last equality, recall that $d \geq (2^n - 1)/(r+1)$ and therefore $2^n = O(d)$. Combining this with (17), we conclude that $s_1 < d/r^{\delta_2}$, for all large enough d. This proves Claim 18. □

Since Claim 18 contradicts Claim 17, we conclude that VP $\neq$ VNP, as claimed in Theorem 3. □

Remarks. 1. We can consider $f'_d(x) := \sum_{i=0}^{d} 2^{i^2} x^i$ and redefine P_n as above. Consider the polynomial $\tilde{P}_n(\boldsymbol{x},\boldsymbol{y})$ defined on $3n$ variables $\boldsymbol{x} = (x_1,\ldots,x_n)$, $\boldsymbol{y} = (y_1,\ldots,y_{2n})$ by $\tilde{P}_n(\boldsymbol{x},\boldsymbol{y}) := \sum_{i=0}^{2^n-1} \phi(i)\cdot \boldsymbol{y}^{\mathrm{bin}(i^2)} \cdot \boldsymbol{x}^{\mathrm{bin}(i)}$, where $\phi(i) := 1$ for all $0 \leq i \leq d$, and 0 otherwise. Note that, substituting $y_j = 2^{2^{j-1}}$ for all $j \in [2n]$ in $\tilde{P}_n$, we get P_n.

We also see: $\tilde{P}_n(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^{2n}-1} \phi(i,j)\cdot \boldsymbol{x}^{\mathrm{bin}(i)}\cdot \boldsymbol{y}^{\mathrm{bin}(j)}$, such that $\phi(i,j) := 1$ when $j = i^2$ with $0 \leq i \leq d$, and 0 otherwise. Note that bit-size of the exponent vector (i.e. sum of bit-size of each co-ordinate) in $\boldsymbol{x}$ and $\boldsymbol{y}$ is $O(n)$. Given $\mathrm{bin}(i)$ and $\mathrm{bin}(j)$, one can easily calculate whether $j = i^2$ or not in $O(n^2)$ time. Hence, $\phi \in$ FP. As, FP $\subset$ #P/poly, therefore by Valiant's criterion, we have $\{\tilde{P}_n\}_n \in$ VNP. As VNP is closed under substitution, we get $\{P_n\}_n \in$ VNP as well!

The hardness part for $\{P_n\}_n$ follows similarly as in the above proof. Thus, we do not need GRH for this particular polynomial!

2. The same proof works for $\prod_{i\in[d]}(x \pm i)$ as for $(x+1)^d$. The hardness part does not change. The only non-trivial part is to show that $\{P_n\}_n \in$ VP, assuming GRH and VP $=$ VNP. The polynomial family $\prod_{i\in[d]}(x \pm i)$ is CH-explicit ([Bür09, Cor.3.12]) and the rest follows similarly.

3.4 Conjecture C1 to matrix rigidity: Proof of Theorem 4

We argue via *linear circuits* which we define first. An arithmetic circuit is called *linear* if it uses only addition gates and multiplications by scalars. As a graph, the nodes of a linear circuits are either input nodes or addition nodes, and the edges are labeled by scalars. If an edge from u to v is labeled by $c \in \mathbb{F}$, then the output of u is multiplied by c and then given as input to v.

Linear circuits can compute linear or affine functions (see [KV19, Sec.1.2]). We give some examples.

1. Let $\boldsymbol{a} = (a_1,\ldots,a_n)$ be a vector. Consider $\boldsymbol{a}$ as a linear function $\mathbb{F}^n \to \mathbb{F}$. It can be computed by a linear circuit of depth 1 with n inputs and one addition-gate as output gate. The edge from the i-th input is labeled by a_i. The size of the circuit is n. However, we omit edges labeled by 0. Hence, the size of the circuit is actually $\mathrm{sp}(\boldsymbol{a}) \leq n$, the sparsity of $\boldsymbol{a}$.

   Similarly, we consider an $n\times n$ matrix A as a linear transformation $\mathbb{F}^n \to \mathbb{F}^n$. For each row vector of A we get a linear circuit as described above. Hence we represent A by circuit of depth 1 with n output gates and size $\mathrm{sp}(A) \leq n^2$.

2. The model gets already interesting for linear circuits of depth 2. Suppose $A = BC$, where B is a $n \times r$ matrix and C is a $r \times n$ matrix. Then we can take the depth-1 circuit for C at the bottom as in item 1 and combine it with the depth-1 circuit for B on top. The resulting depth-2 circuit is *layered*: all edges go either from the bottom to the middle layer, or from the middle to the top layer. The size of the circuit is $\mathrm{sp}(B) + \mathrm{sp}(C) \le 2rn$.

   In particular, there is a representation $A = BC$ with $r = \mathrm{rank}(A)$. Hence the rank of A is involved in the circuit size bound for A. Also note that r is bounded by the size of the circuit because be omit all zero-edges.

   Note that any layered linear circuit of depth 2 in turn gives a factorization of A as a product of 2 matrices, $A = BC$, where the top edges define B and the bottom edges C.

3. Let $A = BC + D$, where B, C are as above and D is a $n \times n$ matrix. The we can represent A by a depth-2 circuit for BC as in item 2 plus edges from the inputs directly to the output nodes to represent D as in item 1. The resulting circuit has depth 2 and size $\mathrm{sp}(B) + \mathrm{sp}(C) + \mathrm{sp}(D) \le 2rn + n^2$, but it would not be layered. We can transform it into a layered circuit by writing A as $A = BC + ID$, where I is the $n \times n$ identity matrix. Then we get a depth-2 circuit for ID similar to BC and can combine the two circuits into one. The size increases by $\le n$ edges for I.

4. Now consider matrix A that is not (r,s)-rigid, for some r, s. Hence, we can write A as $A = R + S$, where $\mathrm{rank}(R) = r$ and $\mathrm{sp}(S) = s$. Then R can be written as as $R = BC$, where B is a $n \times r$ matrix and C is a $r \times n$ matrix. From item 3, we have that $A = BC + S$ has a layered linear circuit of depth 2 of size $\le 2rn + s + n$.

Proof of Theorem 4. The assumption of the theorem is that $U_{\mathbb{F}}(f_d, 2, d) \ge d/2^{\delta_2}$, for some $\delta_2 \ge 1$ and for $d =: n^2 - 1 =: 2^{2\ell} - 1 \in I_2$, for some $n, l \in \mathbb{N}$. Define the bivariate polynomial $g_n \in \mathbb{F}[x, y]$ from f_d via the inverse Kronecker map, $g_n(x, y) = \psi_{2,d}(f_d)$. Recall that g_n has individual degree $\le n - 1$. Equivalently, we can write $g_n(x, y) = \psi_{2,d}(f_d)$. By definition of the Kronecker map, that means $g_n(x, x^n) = f_d(x)$.

Let $g_n(x, y) = \sum_{1 \le i,j \le n} a_{i,j}\, x^{i-1} y^{j-1}$. By the definition of f_d, we have $a_{i,j} = \binom{d}{i-1+(j-1)n}$. Define the $n \times n$ matrix $A_n = (a_{i,j})_{1 \le i,j \le n}$ and vectors

$$[x]_n = \begin{pmatrix} 1 & x & \cdots & x^{n-1} \end{pmatrix}$$
$$[y]_n = \begin{pmatrix} 1 & y & \cdots & y^{n-1} \end{pmatrix}$$

Then we have $g_n(x, y) = [x]_n A_n [y]_n^T$. Next we show a lower bound on the linear circuit size of A_n.

Claim 19. C1 *(with* $\delta_1 = 1, r = 2$*)* $\implies$ *any layered linear circuit of depth 2 that computes* A_n *has size* $> d/2^{\delta_2+1}$.

Proof of Claim 19. Assume that the claim is false. Then we can write $A_n = BC$, where $B \in \mathbb{F}^{n \times t}$, $C \in \mathbb{F}^{t \times n}$, such that $t \le \mathrm{sp}(B) \cup \mathrm{sp}(C) \le d/2^{\delta_2+1}$.

Denote

$$[x]_n\, B \;=\; \begin{pmatrix} \ell_1(x) & \ell_2(x) & \cdots & l_t(x) \end{pmatrix} \;\text{ and }\; C\,[y]_n^T \;=\; \begin{pmatrix} \tilde{\ell}_1(y) & \tilde{\ell}_2(y) & \cdots & \tilde{\ell}_t(y) \end{pmatrix}^T .$$

Then

$$g_n(x, y) = [x]_n\, A_n [y]_n^T = [x]_n\, B\, C [y]_n^T = \sum_{i=1}^{t} \ell_i(x)\, \tilde{\ell}_i(y)\,.$$

Since $\mathrm{sp}(B) \cup \mathrm{sp}(C) \le d/2^{\delta_2+1}$, we have $\sum_{i=1}^{t}(|\ell_i|_1 + |\tilde{\ell}_i|_1) \le d/2^{\delta_2+1}$. In particular, the support-union size $|\bigcup_{i=1}^{t} (\mathrm{supp}(\ell_i) \cup \mathrm{supp}(\tilde{\ell}_i))| \le d/2^{\delta_2+1}$. Substituting $y = x^n$ we get

$$f_d(x) \;=\; g(x, x^n) \;=\; \sum_{i=1}^{t} \ell_i(x)\,\tilde{\ell}_i(x^n) \;=\; \sum_{i=1}^{t}\left(\frac{\ell_i(x)+\tilde{\ell}_i(x^n)}{2}\right)^2 - \sum_{i=1}^{t}\left(\frac{\ell_i(x)-\tilde{\ell}_i(x^n)}{2}\right)^2 .$$

Thus, we have a representation of f_d as $\le 2t \le d/2^{\delta_2}$ sum of squares. Note that addition and subtraction does not increase the support size and thus, we have for the support-union size

$$|\bigcup_{i=1}^{t} \mathrm{supp}(\ell_i \pm \tilde{\ell}_i)| \le |\bigcup_{i=1}^{t} \mathrm{supp}(\ell_i) \cup \mathrm{supp}(\tilde{\ell}_i)| \le d/2^{\delta_2+1} . \tag{18}$$

On the other hand, if Conjecture C1 is true, then we have

$$|\bigcup_{i=1}^{t} \mathrm{supp}(\ell_i \pm \tilde{\ell}_i)| \ge d/2^{\delta_2} . \tag{19}$$

Hence, by (18) and (19), we have a contradiction. This proves Claim 19. □

We want to show that A_n is $(n/2^{\delta_2+3}, n^{1+\delta})$-rigid, for any $\delta < 1$. For the sake of contradiction, assume that this is false. Then there is a $\delta < 1$ and a decomposition $A_n = R + S$, where $\mathrm{rank}(R) = r = n/2^{\delta_2+3}$ and $\mathrm{sp}(S) = s = n^{1+\delta}$. By item 4 from above, A_n has a layered linear circuit C_n of depth 2 of size

$$\mathrm{size}(C_n) \;\le\; 2rn + s + n \;\le\; \frac{2n^2}{2^{\delta_2+3}} + 2n^{1+\delta} . \tag{20}$$

Recall that δ_2 is a constant. Hence, for large enough n, we have $2n^{1+\delta} \le \frac{2n^2-4}{2^{\delta_2+3}}$. Then we can continue the inequalities in (20) by

$$\mathrm{size}(C_n) \;\le\; \frac{4n^2-4}{2^{\delta_2+3}} \;=\; \frac{d}{2^{\delta_2+1}} . \tag{21}$$

For the last equation, recall that $d = n^2 - 1$. The bound in (21) contradicts Claim 19. Therefore we conclude that A_n is $(n/2^{\delta_2+3}, n^{1+\delta})$-rigid. □

3.5 Other models and measures

Kumar and Volk [KV19] showed a strong connection between matrix rigidity and depth-2 linear circuit lower bound. They argued (similarly done in [Pud94] in a different language) that depth-2 $\Omega(n^2)$ lower bound for an explicit matrix is *necessary* and *sufficient* for proving *super*-linear lower bound for general $O(\log n)$-depth circuits (or matrix rigidity).

Symmetric depth-2 circuit. Over $\mathbb{R}$, it is a circuit of the form $B^T \cdot B$, for some $B \in \mathbb{R}^{m\times n}$. [Over $\mathbb{C}$, one should take the conjugate-transpose B^* instead of B^T.] Symmetric circuits are a natural computational model for computing a positive semi-definite (*PSD*) matrix.

Invertible depth-2 circuit. It is a circuit $B \cdot C$, where at least one of the matrices B, C is invertible. We stress that invertible circuits can compute non-invertible matrices. Invertible circuits generalize many of the common matrix decompositions, such as QR decomposition, eigen decomposition, singular value decomposition (SVD), and LUP decomposition.

[KV19, Thms.1.3 & 1.5] also prove asymptotically optimal lower bounds for both the models.

Theorem 20. [KV19] *There exists an explicit family of real $n \times n$ PSD matrices $\{A_n\}_{n\in\mathbb{N}}$ such that every symmetric circuit (resp. invertible circuits) computing A_n (over $\mathbb{R}$) has size $\Omega(n^2)$.*

We present a simple, *alternative* proof of this theorem using SOS representation of f_d over $\mathbb{R}$. For details, see Theorems 33 and 36.

Sparsity-sum measure. We define another *natural* complexity measure–*sparsity-sum*, $S_{\mathbb{F}}(f,r,s)$, with some parameters r,s for a univariate polynomial $f(x)$. It is the minimal sum of sparsity of ℓ_i's in sum-of-rth-powers: $f = \sum_{i\in[s]} c_i \ell_i^r$. Formally,

$$S_{\mathbb{F}}(f,r,s) \;:=\; \min\left(\sum_{i\in[s]} |\ell_i|_1 \;:\; f \;=\; \sum_{i\in[s]} c_i\,\ell_i^r\,,\ \text{where}\ \ c_i \in \mathbb{F}\,,\ \forall\, i \in [s]\right).$$

If such representation does not exist, then it is defined to be ∞. Note that $U_{\mathbb{F}}(\cdot) \le S_{\mathbb{F}}(\cdot)$.

We note that our Theorems 2–4 hold for the 'larger' measure $S_{\mathbb{F}}$ (see Theorem 41). However, it is *not* clear whether some lower bound on $S_{\mathbb{F}}$ could give an efficient HSG for VP-circuits.

In Theorem 40, we prove a lower bound of $S_{\mathbb{R}}(f,r,s) \ge \Omega(d^{1/r^{\log(4/3)}})$, for a bivariate d-sparse polynomial $f(x,y)$ of *individual* degree d. This is better than the trivial lower bound of $\Omega(d^{1/r})$, as $\log_2(4/3) < 1$.

4 Conclusion

Since our Conjecture C1 and its underlying framework is new, many lines of investigation have opened up. Here are some immediate questions of interest.

1. Is Conjecture C1 true for a 'generic' polynomial f with rational coefficients?
2. Prove, or disprove, Conjecture C1 for *constants* r,s. In particular, can we say that $U_{\mathbb{R}}((x+1)^d, 2, s) \ge \Omega(d)$, for all constants s? What about $S_{\mathbb{R}}(\cdot)$?
3. Prove $U_{\mathbb{Z}}((x+1)^d, r, s) \ge \Omega(d/r)$, for *all* large enough d (i.e. for the ones outside I_r too).
4. Can we show that the connection between conjecture and PIT $\in$ P holds, for smaller prime-powers $r < 25$. In particular, does SOS lower bounds solve PIT, or VP $\neq$ VNP ?
5. Can we remove the GRH assumption for the polynomial $(x+1)^d$ (in Theorem 3)?

Acknowledgments. We thank Nikhil Balaji and Abhiroop Sanyal for the useful discussions in the initial phase of the project. P.D. thanks CSE, IIT Kanpur for the hospitality, and acknowledges the support of Google PhD Fellowship. N.S. thanks the funding support from DST (DST/SJF/MSA-01/2013-14) and N. Rama Rao Chair. T.T. thanks DFG for the funding (grant TH 472/5-1), and CSE, IIT Kanpur for the hospitality.

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009. 7

[AC19] Josh Alman and Lijie Chen. Efficient construction of rigid matrices using an NP oracle. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1034–1055. IEEE, 2019. 5

[AGS19] Manindra Agrawal, Sumanta Ghosh, and Nitin Saxena. Bootstrapping variables in algebraic circuits. *Proceedings of the National Academy of Sciences*, 116(17):8107–8118, 2019. (Earlier in STOC'18). 2, 4

[AJMV98] Eric Allender, Jia Jiao, Meena Mahajan, and V. Vinay. Non-commutative arithmetic circuits: Depth reduction and size lower bounds. *Theor. Comput. Sci.*, 209(1-2):47–86, 1998. 8, 28

[AV08] Manindra Agrawal and V Vinay. Arithmetic circuits: A chasm at depth four. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 67–75. IEEE, 2008. 2, 3, 4

[BCS13] Peter Bürgisser, Michael Clausen, and Amin Shokrollahi. *Algebraic Complexity Theory*, volume 315. Springer Science & Business Media, 2013. 1, 2

[Bel58] Edouard Grigor'evich Belaga. Some problems involved in the calculation of polynomials. In *Doklady Akademii Nauk*, volume 123, pages 775–777. Russian Academy of Sciences, 1958. 2

[BM16] Boaz Barak and Ankur Moitra. Noisy tensor completion via the Sum-of-squares Hierarchy. In *Conference on Learning Theory*, pages 417–445, 2016. 2

[BT15] Grigoriy Blekherman and Zach Teitler. On maximum, typical and generic ranks. *Mathematische Annalen*, 362(3-4):1021–1031, 2015. 2

[Bür00] Peter Bürgisser. Cook's versus Valiant's hypothesis. *Theoretical Computer Science*, 235(1):71–88, 2000. 8

[Bür09] Peter Bürgisser. On defining integers and proving arithmetic circuit lower bounds. *Computational Complexity*, 18(1):81–103, 2009. 2, 3, 15, 17

[Bür13] Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*, volume 7. Springer Science & Business Media, 2013. 2

[CCG12] Enrico Carlini, Maria Virginia Catalisano, and Anthony V Geramita. The solution to the waring problem for monomials and the sum of coprime monomials. *Journal of algebra*, 370:5–14, 2012. 2

[DGW19] Zeev Dvir, Alexander Golovnev, and Omri Weinstein. Static data structure lower bounds imply rigidity. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 967–978, 2019. 5

[Dix64] John D Dixon. Another proof of Lagrange's four square theorem. *The American Mathematical Monthly*, 71(3):286–288, 1964. 2

[DL78] Richard A. Demillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193 – 195, 1978. 2

[EPRS08] Friedrich Eisenbrand, János Pach, Thomas Rothvoß, and Nir B Sopher. Convexly independent subsets of the Minkowski sum of planar point sets. *The electronic journal of combinatorics*, 15(1):8, 2008. 33

[Fis94] Ismor Fischer. Sums of like Powers of Multivariate Linear Forms. *Mathematics Magazine*, 67(1):59–61, 1994. 8

[FOS12] Ralf Fröberg, Giorgio Ottaviani, and Boris Shapiro. On the waring problem for polynomial rings. *Proceedings of the National Academy of Sciences*, 109(15):5600–5602, 2012. 2

[GKKS13] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi. Arithmetic circuits: A chasm at depth three. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 578–587. IEEE, 2013. 2

[GKSS19] Zeyu Guo, Mrinal Kumar, Ramprasad Saptharishi, and Noam Solomon. Derandomization from algebraic hardness: Treading the borders. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 147–157, 2019. 2, 4, 5, 10

[GMK17] Ignacio Garcia-Marco and Pascal Koiran. Lower bounds by Birkhoff interpolation. *Journal of Complexity*, 39:38–50, 2017. 3

[GMQ16] Joshua A. Grochow, Ketan D. Mulmuley, and Youming Qiao. Boundaries of VP and VNP. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:14, 2016. 2

[Gro15] Joshua A Grochow. Unifying known lower bounds via geometric complexity theory. *Computational Complexity*, 24(2):393–475, 2015. 2

[HS80] Joos Heintz and Malte Sieveking. Lower bounds for polynomials with algebraic coefficients. *Theoretical Computer Science*, 11(3):321–330, 1980. 2

[HWY11] Pavel Hrubeš, Avi Wigderson, and Amir Yehudayoff. Non-commutative circuits and the sum-of-squares problem. *Journal of the American Mathematical Society*, 24(3):871–898, 2011. 5

[Kem12] Aubrey Kempner. Bemerkungen zum Waringschen problem. *Mathematische Annalen*, 72(3):387–399, 1912. 2

[KI03] Valentine Kabanets and Russell Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 355–364. ACM, 2003. 2, 4

[KKPS15] Neeraj Kayal, Pascal Koiran, Timothée Pecatte, and Chandan Saha. Lower bounds for sums of powers of low degree univariates. In *International Colloquium on Automata, Languages, and Programming*, pages 810–821. Springer, 2015. 3

[Koi11] Pascal Koiran. Shallow circuits with high-powered inputs. In *Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011. Proceedings*, pages 309–320, 2011. 2, 3

[Koi12] Pascal Koiran. Arithmetic circuits: The chasm at depth four gets wider. *Theoretical Computer Science*, 448:56–65, 2012. 2, 3, 4

[KP11] Pascal Koiran and Sylvain Perifel. Interpolation in Valiant's theory. *computational complexity*, 20(1):1–20, 2011. 15

[KPGM18] Pascal Koiran, Timothée Pecatte, and Ignacio García-Marco. On the linear independence of shifted powers. *Journal of Complexity*, 45:67–82, 2018. 3

[KPTT15] Pascal Koiran, Natacha Portier, Sébastien Tavenas, and Stéphan Thomassé. A τ-conjecture for newton polygons. *Foundations of computational mathematics*, 15(1):185–197, 2015. 2, 33

[Kro82] Leopold Kronecker. Grundzüge einer arithmetischen theorie der algebraischen grössen.(abdruck einer festschrift zu herrn ee kummers doctor-jubiläum, 10. september 1881.). *Journal für die reine und angewandte Mathematik*, 92:1–122, 1882. 9

[KS19] Mrinal Kumar and Ramprasad Saptharishi. Hardness-randomness tradeoffs for algebraic computation. *Bulletin of EATCS*, 1(129), 2019. 2

[KSS19] Mrinal Kumar, Ramprasad Saptharishi, and Noam Solomon. Derandomization from Algebraic Hardness: Treading the Borders. `https://arxiv.org/pdf/1905.00091v1.pdf`, 2019. 29

[KST19] Mrinal Kumar, Ramprasad Saptharishi, and Anamay Tengse. Near-optimal Bootstrapping of Hitting Sets for Algebraic Circuits. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 639–646, 2019. 2

[KV19] Mrinal Kumar and Ben Lee Volk. Lower Bounds for Matrix Factorization. `https://arxiv.org/pdf/1904.01182.pdf`, 2019. 17, 19, 20, 31, 32

[Lan13] Serge Lang. *Algebraic number theory*, volume 110. Springer Science & Business Media, 2013. 4

[Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009. 2

[Lok09] Satyanarayana V Lokam. Complexity lower bounds using linear algebra. *Foundations and Trends® in Theoretical Computer Science*, 4(1–2):1–155, 2009. 5

[Luc78] Edouard Lucas. Théorie des fonctions numériques simplement périodiques. *American Journal of Mathematics*, pages 289–321, 1878. 13

[Mah14] Meena Mahajan. Algebraic Complexity Classes. In *Perspectives in Computational Complexity*, pages 51–75. Springer, 2014. 1

[Mot55] T.S. Motzkin. Evaluation of polynomials and evaluation of rational functions. *Bull. Amer. Math. Soc*, 61(163):10, 1955. 2

[Muk16] Partha Mukhopadhyay. Depth-4 identity testing and Noether's normalization lemma. In *International Computer Science Symposium in Russia*, pages 309–323. Springer, 2016. 2

[Mul12] Ketan D. Mulmuley. Geometric complexity theory V: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's normalization lemma. In *FOCS*, pages 629–638, 2012. 2

[Mul17] Ketan Mulmuley. Geometric complexity theory V: Efficient algorithms for Noether normalization. *Journal of the American Mathematical Society*, 30(1):225–309, 2017. 2

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of computer and System Sciences*, 49(2):149–167, 1994. 1

[Ore22] Øystein Ore. Über höhere kongruenzen. *Norsk Mat. Forenings Skrifter*, 1(7):15, 1922. 2

[Ost75] Alexander M Ostrowski. On multiplication and factorization of polynomials, I. Lexicographic orderings and extreme aggregates of terms. *aequationes mathematicae*, 13(3):201–228, 1975. 33

[Pud94] Pavel Pudlak. Communication in bounded depth circuits. *Combinatorica*, 14(2):203–216, 1994. 19

[Sax09] Nitin Saxena. Progress on Polynomial Identity testing. *Bulletin of the EATCS*, 99:49–79, 2009. 2

[Sax14] Nitin Saxena. Progress on Polynomial Identity Testing - II. *Perspectives in Computational Complexity*, 26:131–146, 2014. 2

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, October 1980. 2

[SS+95] Michael Shub, Steve Smale, et al. On the intractability of Hilbert's Nullstellensatz and an algebraic version of " NP$\neq$ P?". *Duke Mathematical Journal*, 81(1):47–54, 1995. 2

[Str74] Volker Strassen. Polynomials with rational coefficients which are hard to compute. *SIAM Journal on Computing*, 3(2):128–149, 1974. 2

[SY10] Amir Shpilka and Amir Yehudayoff. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends® in Theoretical Computer Science*, 5(3–4):207–388, 2010. 1, 2, 5

[Tav15] Sébastien Tavenas. Improved bounds for reduction to depth 4 and depth 3. *Information and Computation*, 240:2–11, 2015. 2, 3, 4

[Val77] Leslie G Valiant. Graph-theoretic arguments in low-level complexity. In *International Symposium on Mathematical Foundations of Computer Science*, pages 162–176. Springer, 1977. 5

[Val79a] Leslie G. Valiant. Completeness classes in algebra. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 249–261, 1979. 1

[Val79b] Leslie G Valiant. Completeness classes in algebra. In *Proceedings of the 11th Annual ACM symposium on Theory of computing*, pages 249–261. ACM, 1979. 8

[VSBR83] Leslie G. Valiant, Sven Skyum, Stuart Berkowitz, and Charles Rackoff. Fast parallel computation of polynomials using few processors. *SIAM Journal on Computing*, 12(4):641–644, 1983. 5, 8, 28

[Wag86] Klaus W Wagner. The complexity of combinatorial problems with succinct input representation. *Acta informatica*, 23(3):325–356, 1986. 7

[Wig17] Avi Wigderson. Low-depth arithmetic circuits: technical perspective. *Communications of the ACM*, 60(6):91–92, 2017. 2

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Computation*, EUROSAM '79, pages 216–226, 1979. 2

A Bounds for sum of constant-powers: Details for Section 1.1

By Lemma 9, we already showed an upper bound for univariate polynomials represented as sum of *few*, namely $(r+1)$ many r-th powers of polynomials over $\mathbb{F}$ of characteristic 0 or large, thus showing that it is a complete model.

A.1 $(x+1)^d$ as sum of $r+1$ many r-th powers

Using Lemma 9, we show an upper bound on $U_{\mathbb{F}}((x+1)^d, r, r+1)$.

Lemma 21. *For any $r \leq d \in \mathbb{N}$, we have $U_{\mathbb{F}}\left((x+1)^d, r, r+1)\right) \leq d/r + r$.*

Proof. Let $d = rk + t$, where $t = d \bmod r$ and $k = \lfloor d/r \rfloor$. Then, from Lemma 9, it follows that there exist $c_i, \lambda_i \in \mathbb{F}$ such that

$$\begin{aligned}
(x+1)^d &= \left((x+1)^k\right)^r (x+1)^t \\
&= \left((x+1)^k\right)^r \sum_{i=0}^{r} c_i \left((x+1)^t + \lambda_i\right)^r \\
&= \sum_{i=0}^{r} c_i \left((x+1)^{t+k} + \lambda_i (x+1)^k\right)^r \ =: \ \sum_{i=0}^{r} c_i\, \ell_i^r
\end{aligned}$$

where $\ell_i := (x+1)^{t+k} + \lambda_i (x+1)^k$. Note that $|\bigcup_{i=0}^{r} \mathrm{supp}(\ell_i)| \leq t + k + 1 \leq d/r + r$. $\square$

A.2 Sum of powers of small support-union

We give a second way how a univariate polynomials can be represented as sum of r-th powers of polynomials. It is a bit more complicated than the one from Lemma 9.

Here we use the notion of sumset. In additive combinatorics, the *sumset*, also called the *Minkowski sum* of two subsets A and B of an abelian group G is defined to be the set of all sums of an element from A with an element from B,

$$A + B \ = \ \{\, a + b \mid a \in A,\ b \in B \,\}\,.$$

The *n-fold iterated sumset* of A is $nA = A + \ \cdots\ + A$, where there are n summands.

We want a *small support-union* representation of a d-degree polynomial f as a sum of r-th powers, where r is constant. We consider a *small* B such that rB covers $\{0, 1, \ldots, d\}$. Let t be the *unique* non-negative integer such that $(t-1)^r < d + 1 \leq t^r$. Define the set B as

$$B \ = \ \{\, a_i\, t^k \mid 0 \leq a_i \leq t-1,\ 0 \leq k \leq r-1 \,\}\,.$$

So $|B| = rt = O(d^{1/r})$. Let $k \in \{0, 1, \ldots, d\}$. The base-$t$ representation of k is a sum of at most elements from B. Hence, $\{0, 1, \ldots, d\} \subseteq rB$.

The largest element in B is $m = (t-1)t^{r-1}$. Note that, for any $\epsilon > 0$, we have $t < (1 + \epsilon)(d+1)^{1/r}$, for all large enough d. Thus, for *any* constant $c > 1$ and large enough d, we have $m < c(d+1)$. Therefore, the largest element in rB is at most $mr < cr(d+1) = O(d)$.

Lemma 22. *Let $\mathbb{F}$ be a field of characteristic 0 or large. For any $f(x) \in \mathbb{F}[x]$ of degree d, there exist $\ell_i \in \mathbb{F}[x]$ with $\mathrm{supp}(\ell_i) \subseteq B$ and $c_i \in \mathbb{F}$, for $i = 0, 1, \ldots, mr$, such that $f(x) \ = \ \sum_{i=0}^{mr} c_i\, \ell_i^r$.*

Proof. Consider $\ell_i(z_i, x) = \sum_{j\in B} z_{ij}x^j$, for distinct indeterminates z_{ij}, for all i, j. Surely, $\deg_x(\ell_i) = m$. There exists $mr + 1$ many degree-r polynomials Q_j over $|B| = rt$ many variables, such that

$$\ell_i(z_i, x)^r \;=\; \sum_{j=0}^{mr} Q_j(z_i)\, x^j \qquad \forall i \in [mr]\,.$$

Note that from any monomial in Q_j we could recover j uniquely. Thus, we could conclude that $Q_j(z_i)$ $(0 \le j \le mr)$ are $\mathbb{F}$-linearly independent.

Suppose $f(x) = \sum_{i=0}^{d} f_i\, x^i$. Define $\tilde{f} \in \mathbb{F}^{mr+1}$ and $A \in \mathbb{F}[z]^{(mr+1)\times(mr+1)}$ as

$$\tilde{f} = \begin{pmatrix} f_0 & f_1 & \cdots & f_d & 0 & \cdots & 0 \end{pmatrix},\; A = \begin{pmatrix} Q_0(z_0) & Q_1(z_0) & \cdots & Q_{mr}(z_0) \\ Q_0(z_1) & Q_1(z_1) & \cdots & Q_{mr}(z_1) \\ \vdots & \vdots & \cdots & \vdots \\ Q_0(z_{mr}) & Q_1(z_{mr}) & \cdots & Q_{mr}(z_{mr}) \end{pmatrix}.$$

We want to find $\boldsymbol{c} = \begin{pmatrix} c_0 & c_1 & \cdots & c_{mr} \end{pmatrix} \in \mathbb{F}^{mr+1}$ and $\boldsymbol{\alpha} = (\alpha_{ij})_{i,j}$ such that

$$\sum_{i=0}^{mr} c_i\, \ell_i(\boldsymbol{\alpha}, x)^r \;=\; \sum_{i=0}^{d} f_i\, x^i \iff \boldsymbol{c}\cdot A|_{z=\boldsymbol{\alpha}} \cdot \begin{pmatrix} 1 \\ x \\ \vdots \\ x^{mr} \end{pmatrix} \;=\; \tilde{f}\cdot \begin{pmatrix} 1 \\ x \\ \vdots \\ x^{mr} \end{pmatrix} \iff \boldsymbol{c}\cdot A|_{z=\boldsymbol{\alpha}} = \tilde{f}\,.$$

As the z_i's are distinct variables, the first column of A consists of different variables at each coordinate. Moreover, the first row of A contains $\mathbb{F}$-linearly independent Q_j's. Thus, for *random* $\alpha_{ij} \in \mathbb{F}$, matrix $A|_{z=\boldsymbol{\alpha}}$ has *full* rank over $\mathbb{F}$. Fix such an $\boldsymbol{\alpha}$. This fixes $\boldsymbol{c} = \tilde{f}\cdot (A|_{z=\boldsymbol{\alpha}})^{-1}$.

From the above construction, it follows that $f(x) \;=\; \sum_{i=0}^{mr} c_i\, \ell_i(\boldsymbol{\alpha}, x)^r$. ☐

Thus, for any d-degree f, $U_{\mathbb{F}}(f, r, s := mr+1) \le O(d^{1/r})$. As seen before, $mr = \Theta(d)$; when $s \ge c\cdot(d+1)$ where $c > r$, we have a small base representation for large enough d, as mr can be made smaller than any constant ($> r$) multiple of $d+1$. It is unclear, though, whether even for $s \le d$, such a small support-union representation exists.

Remarks.

1. The above calculation does *not* give small sparsity-sum representation of f, as the top fan-in is already $\Omega(d)$.
2. Both the above representations (small s resp. small support-union) crucially require a *field* $\mathbb{F}$. E.g. they do not exist for f_d over the ring $\mathbb{Z}$ by Theorem 2.

A.3 $(x+1)^d$ as sum of two r-th powers

We show a strong lower bound of $\Omega(d/r)$ for $f_d(x) := (x+1)^d$ when written as sum of two r-th powers. W.l.o.g., we consider $\mathbb{F}$ algebraically closed, as $U_{\overline{\mathbb{F}}}(\cdot) \le U_{\mathbb{F}}(\cdot)$. Note that, $c_1\cdot \ell_1^r + c_2\cdot \ell_2^r = \tilde{\ell}_1^r - \tilde{\ell}_2^r$ where $\tilde{\ell}_1 = c_1^{1/r}\cdot \ell_1$ and $\tilde{\ell}_2 = (-c_2)^{1/r}\cdot \ell_2$. Also, $|\bigcup_{i=1}^{2}\mathrm{supp}(\ell_i)| = |\bigcup_{i=1}^{2}\mathrm{supp}(\tilde{\ell}_i)|$. Thus, it suffices to prove the bounds when f_d is written as $\ell_1^r - \ell_2^r$. Before that, we prove the following.

Lemma 23. *For a fixed $d \ge 1$ and $r \ge 3$, if $(x+1)^d = \ell_1^r - \ell_2^r$, for some $\ell_i \in \mathbb{F}[x]$, then ℓ_1 and ℓ_2 must share a non-trivial gcd.*

Proof. Suppose, $\gcd(\ell_1, \ell_2) = 1$. Note that, $\ell_1^r - \ell_2^r$ has the following factorization over $\mathbb{F}[x]$,

$$(x+1)^d \;=\; (\ell_1 - \ell_2)(\ell_1 - \zeta_r \ell_2) \ldots (\ell_1 - \zeta_r^{r-1} \ell_2)$$

where ζ_r is a primitive r-th root of unity. If $(x+1) \mid (\ell_1 - \zeta_r^i \ell_2)$ and $(\ell_1 - \zeta_r^j \ell_2)$, for $i \neq j$, then subtraction would imply: $(x+1) \mid \ell_1, \ell_2$. This contradicts our assumption; hence, there must exist i: $\ell_1 - \zeta_r^i \ell_2 \;=\; c \cdot (x+1)^d$. In particular, it means: $\ell_1 - \zeta_r^j \ell_2$ is constant, for all $j \neq i$. Subtracting two such equations immediately gives us: ℓ_1, ℓ_2 are constants; a contradiction again as $d \geq 1$. $\square$

Corollary 24. *For $3 \leq r \leq d$: $(x+1)^d = \ell_1^r - \ell_2^r$ iff $r | d$. In that case, $\exists \alpha_1, \alpha_2 \in \mathbb{F}$ such that $\ell_i = \alpha_i \cdot (x+1)^{d/r}$.*

Proof. By Lemma 23, $\gcd(\ell_1, \ell_2) =: p(x)$ is non-constant. Therefore, $p^r \mid (x+1)^d$; implying that $p(x)$ is a power of $x+1$. After dividing out, we can again apply the lemma; eventually, we deduce $r|d$. It also implies: $\ell_i = \alpha_i \cdot (x+1)^{d/r}$, for some $\alpha_i \in \mathbb{F}$. $\square$

Theorem 25. *For any $d \geq 1$ and any $r \geq 2$, we have*

$$U_{\mathbb{F}}(f_d, r, 2) = \begin{cases} \lceil d/r \rceil + 1 & \text{if } r \mid d \text{ or } r = 2\,, \\ \infty & \text{otherwise.} \end{cases}$$

Proof. We prove this in two separate cases.

Case I: For $r \geq 3$, the above corollary implies that $r|d$ and that support-union is $d/r + 1$.

Case II: Consider $r = 2$. In this case, we have $(x+1)^d = (\ell_1 + \ell_2) \cdot (\ell_1 - \ell_2)$. Thus, there exists $c_1, c_2 \in \mathbb{F}$ and $0 \leq t \leq d$ such that $(\ell_1 + \ell_2) = c_1 (x+1)^t$ and $(\ell_1 - \ell_2) = c_2 (x+1)^{d-t}$.

This implies: $|\bigcup_{i=1}^2 \mathrm{supp}(\ell_i)| \geq \max(d-t+1, t+1) \geq d/2 + 1$. In fact, one can choose $t = \lfloor d/2 \rfloor$; in that case $U_{\mathbb{F}}(\cdot) = \lceil d/2 \rceil + 1$. $\square$

B Hitting-set for $\overline{\mathsf{VP}}$: Details for Section 3.1

Here we study hitting-set for the approximative class $\overline{\mathsf{VP}}$. Before doing that, it is important to recall the meaning of approximation in the algebraic setting.

Definition 26 (Approximative computation)**.** *A circuit C over $\mathbb{F}(\epsilon)[\boldsymbol{x}]$ is said to* approximate *a polynomial $P(\boldsymbol{x})$, if we can write $C(\boldsymbol{x}, \epsilon) = \epsilon^M P(\boldsymbol{x}) + \epsilon^{M+1} Q(\boldsymbol{x}, \epsilon)$, for some polynomial $Q(\boldsymbol{x}, \epsilon) \in \mathbb{F}[\boldsymbol{x}, \epsilon]$ and $M \in \mathbb{N}$. In other words,*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^M} C(\boldsymbol{x}, \epsilon) \;=\; P(\boldsymbol{x})\,.$$

We denote by $\overline{\mathrm{size}}(P)$, the *approximative circuit complexity* of P to be the size of the smallest circuit that approximates P. The class $\overline{\mathsf{VP}}$ contains the families of n-variate polynomials of degree $n^{O(1)}$ over $\mathbb{F}$ of approximative complexity $n^{O(1)}$.

B.1 Tools for $\overline{\mathsf{VP}}$

We point out that the log-depth reduction [VSBR83, AJMV98] works over approximative circuits as well.

Theorem 27. *Suppose $f(\boldsymbol{x}) \in \mathbb{F}[\boldsymbol{x}]$ is a polynomial of degree d which can be approximated by a s size circuit C. Then, there exists a normal-form circuit C' of size $O(s^3 d^6)$ approximating the same f.*

Proof sketch. Suppose f is approximated by a circuit which computes $C(\boldsymbol{x}, \epsilon)$. One can show that each homogeneous part of C (w.r.t. $\boldsymbol{x}$) can be approximated by a circuit of size $s' = O(sd^2)$. Hence, it suffices to depth-reduce homogeneous circuits and show that $O(s'^3)$ size normal-form circuits compute the same polynomial C. $\square$

Kumar et al. [KSS19] proved that the hardness of constant-variate polynomials in the approximative sense, suffices to construct an HSG for $\overline{\mathsf{VP}}$.

Theorem 28. [KSS19, Thm.1.6] *Let $P \in \mathbb{F}[\boldsymbol{x}]$ be a k-variate polynomial of degree d. Suppose $\overline{\mathsf{size}}(P) > sDdn^{10k}$, for parameters n, D, s, then there is a $\mathsf{poly}(s)$-time HSG for any $(n+1)$-variate polynomial $Q(x_0, \ldots, x_n)$ of degree D such that $\overline{\mathsf{size}}(Q) \le s$.*

B.2 Hitting-set for $\overline{\mathsf{VP}}$: Approximative version of Theorem 1

Let field $\mathbb{F}$ be $\mathbb{Q}, \mathbb{Q}_p$ (or their fixed extensions), or a finite field of large characteristic. Let us first formalize Conjecture C1 in the approximative setting. For a ring R, we define *support-union approximative size* $\overline{U}_R(f, r, s)$ as the number of distinct monomials required to *approximate* f as sum-of-r^{th}-powers. In particular, define

$$\overline{U}_R(f,r,s) \;:=\; \min\left(\left|\bigcup_{i=1}^{s}\mathrm{supp}(\ell_i)\right| \;:\; g(x,\epsilon) = \sum_{i=1}^{s} c_i \ell_i^r \;\text{ and }\; \lim_{\epsilon\to 0}\frac{1}{\epsilon^M}\cdot g = f, \text{ for some } M \ge 0\right).$$

Obviously, $\overline{U}_R(\cdot) \le U_R(\cdot)$. We conjecture that even $\overline{U}_{\mathbb{F}}(f_d, r, s)$ is large for $f_d := (x+1)^d$.

Conjecture 2 (C2). *There exist positive constants $\delta_1 \le 1$, $\delta_2 \ge 1$ and a constant prime-power r such that $\overline{U}_{\mathbb{F}}\left(f_d, r, d^{\delta_1}\right) \ge d/r^{\delta_2}$, for all large enough $d \in I_r$.*

Theorem 29. *If Conjecture C2 holds true for an $r \ge 25$, then there is a poly-time HSG for $\overline{\mathsf{VP}}$-circuits.*

Proof sketch. The proof is almost the same as that of Theorem 1. We define P_n similarly (i.e. inverse Kronecker applied on f_d where d was chosen uniquely from an interval based on n). We claim that $\overline{\mathsf{size}}(P_n) > d^{1/\mu}$, where $\mu \ge 3/(\delta_1/r - 7/k)$ (same as in Section 3.1).

Hardness of P_n: We assume that there is a circuit C of size at most $d^{1/\mu}$ computing a polynomial $C(\boldsymbol{x}, \epsilon) \in \mathbb{F}(\epsilon)[\boldsymbol{x}]$, which approximates P_n over large enough $n \in J$, where $J \subseteq \mathbb{N}$ is an infinite subset. Using Theorem 27, there exists a normal-form circuit C' of size at most $s' := d^{3/\mu} \cdot (kn)^6$ approximating P_n. Assume that, $C'(\boldsymbol{x}, \epsilon) =: \epsilon^M \cdot P_n + \epsilon^{M+1} \cdot Q(\boldsymbol{x}, \epsilon)$, for some $M \in \mathbb{N}_{\ge 0}$.

We can depth-reduce C' to depth-4 with some constant t (as done in the proof of Theorem 1) so that $C'(x, \epsilon) = \sum_{i=1}^{\tilde{s}} \tilde{c}_i \cdot \tilde{g}_i^r$, where $\tilde{s} := s_0 \cdot 2^{5^t} \cdot (r+1)$, and each $\tilde{g}_i$ is a k-variate polynomial of degree at most $k \cdot n \cdot 2^{-t}$ over $\mathbb{F}(\epsilon)[\boldsymbol{x}]$. We apply $\phi_{k,n}$ on $C'(x, \epsilon)$. As, $\phi_{k,n} \circ \psi_{k,d} = \mathrm{id}$, over $\mathbb{F}[x]^{\le d}$. Thus,

$$\epsilon^M \cdot f_d + \epsilon^{M+1} \cdot \tilde{Q} \;:=\; (\phi_{k,n} \circ \psi_{k,d})(C') \;=\; \sum_{i=1}^{\tilde{s}} \tilde{c}_i \cdot (\phi_{k,n}(\tilde{g}_i))^r$$

where we have used that $\phi_{k,n}(P_n) = f_d$ and $\phi_{k,n}(Q(\boldsymbol{x}, \epsilon)) = \tilde{Q}(x, \epsilon)$, for some $\tilde{Q} \in \mathbb{F}[x, \epsilon]$. It is important to observe that $|\bigcup_i \mathrm{supp}(\tilde{g}_i)| \le s_1 := \binom{k+k\cdot n\cdot 2^{-t}}{k}$. Since Kronecker map can not increase the support size, therefore $|\bigcup_i \mathrm{supp}((\phi_{k,n}(\tilde{g}_i)))| \le m$. Thus, we must have $\overline{U}_{\mathbb{F}}(f_d, r, \tilde{s}) \le s_1$ from the definition of $\overline{U}_{\mathbb{F}}(\cdot)$.

We can show that $\tilde{s} < d^{\delta_1}$ and $s_1 < d/r^{\delta_2}$, for all large enough n, where $k \geq 17(\delta_2+1)$ and $t \geq 2$ (as shown in Section 3.1). Therefore, we have $\overline{U}_{\mathbb{F}}(f_d, r, d^{\delta_1}) < d/r^{\delta_2}$, over all large $d \in J_r := \{d(n) \in I_r \mid n \in J\} \subseteq I_r$. This contradicts Conjecture C2. Thus, $\overline{\text{size}}(P_n) > d^{1/\mu}$, for a suitable constant μ and all large enough n.

Like in Section 3.1, P_n is explicit and hard; thus Theorem 28 gives us a poly(s)-time HSG for $\overline{\text{size}}$-s degree-s polynomials. $\square$

C SOS with restrictions: Details for Section 3.5

In this section, we will prove two lower bounds of SOS representation (with restriction), and give our alternative proof of Theorem 20.

C.1 Lower bound for symmetric circuits over $\mathbb{R}$: Proof of the first part of Thm.20

We state a lemma from classical mathematics for the study of fewnomials and give a simple proof.

Lemma 30 (Hajós Lemma). *Suppose $f(x) \in \mathbb{C}[x]$ be a univariate polynomial with $t \geq 1$ monomials. Let α be a non-zero root of $f(x)$. Then, the multiplicity of α in f can be at most $t-1$.*

Proof. We will prove this by induction on t. When $t = 1$, $f(x) = a_m x^m$ for some m. It has no non-zero roots and we are trivially done. Assume that, it is true upto t. We want to prove the claim for $t+1$.

Suppose $|f|_1 = t+1$. There exists $m \geq 0$ such that $f(x) = x^m.g(x)$ with $|g|_1 = t+1$ and $g(0) \neq 0$. It suffices to prove the claim for g. Let, α be a non-zero root of $g(x)$. Suppose, $g(x) = (x-\alpha)^s \cdot h(x)$, for some $s \geq 1$ and $h(\alpha) \neq 0$. Observe that, multiplicity of α in g' is $s-1$. As $g(0) \neq 0$, $|g'|_1 = t$. Therefore by induction hypothesis, $s-1 \leq t-1 \implies s \leq t$. Hence, multiplicity of α in g (thus in f) can be at most t. This finishes the induction step. $\square$

Corollary 31. *Suppose $f(x) = (x+\alpha)^t \cdot g(x)$, for some non-zero α and $g(\cdot)$, then we must have $|f|_1 \geq t+1$.*

We prove an important lower bound on SOS representation for a non-zero multiple of $(x+1)^d$; it will be important to prove the first part of Theorem 20.

Lemma 32. *Let $f(x)$ be a non-zero polynomial in $\mathbb{R}[x]$. Suppose, there exist non-zero $\ell_i \in \mathbb{R}[x]$, for $i \in [m]$ such that $(x+1)^d \cdot f(x) = \sum_{i=1}^m \ell_i^2$. Then, $\sum_{i\in[m]} |\ell_i|_1 \geq m \cdot (\lfloor d/2 \rfloor + 1)$.*

Proof. Denote $g(x) := \gcd(\ell_1, \ldots, \ell_m)$. We will prove that $(x+1)^t \mid g(x)$ where $t := \lfloor d/2 \rfloor$. Suppose not, assume that $(x+1)^k || g(x)$ (i.e $(x+1)^{k+1} \nmid g(x)$) such that $k < t$ (and thus $d-2k > 0$). Then, $g(x) = h(x) \cdot (x+1)^k$ where $h(x) \in \mathbb{R}[x]$ with $h(-1) \neq 0$. Define $\tilde{\ell}_i := \ell_i/(x+1)^k$. By assumption, $(x+1) \nmid \gcd(\tilde{\ell}_1, \ldots, \tilde{\ell}_m) =: h(x)$. Thus,

$$\begin{aligned}
\sum_{i=1}^{k} \ell_i(x)^2 \;=\; (x+1)^d \cdot f(x) \;&\implies\; \sum_{i=1}^{m} \tilde{\ell}_i(x)^2 \;=\; (x+1)^{d-2k} \cdot f(x) \\
&\implies\; \sum_{i=1}^{m} \tilde{\ell}_i(-1)^2 \;=\; 0 \\
&\implies\; \tilde{\ell}_i(-1) \;=\; 0\,, \quad \forall i \in [1,m] \\
&\implies\; (x+1) \mid \tilde{\ell}_i(x)\,, \quad \forall i \in [1,m] \\
&\implies\; (x+1) \mid \gcd(\tilde{\ell}_1, \ldots, \tilde{\ell}_m) = h(x)
\end{aligned}$$

which is a contradiction. Thus, $k \geq t$.

This implies, each ℓ_i is non-zero polynomial multiple of $(x+1)^t$. Since Corollary 31 implies that $|\ell_i|_1 \geq t+1$, for all $i \in [m]$; the lemma follows. $\square$

Recall that a *symmetric* depth-2 circuit (over $\mathbb{R}$) is a circuit of the form $B^T \cdot B$ for some $B \in \mathbb{R}^{m \times n}$. We prove the *first* part of Theorem 20.

Theorem 33 (Reproving Thm.1.3 of [KV19])**.** *There exists an explicit family of real $n \times n$ PSD matrices $\{A_n\}_{n \in \mathbb{N}}$ such that every symmetric circuit computing A_n (over $\mathbb{R}$) has size $\Omega(n^2)$.*

Proof. Denote $[x]_n := \begin{bmatrix} 1 & x & \ldots & x^{n-1} \end{bmatrix}$. Denote $k := \lfloor n/2 \rfloor$. Define $g_i(x) := (x+1)^k \cdot x^{\lfloor (i-1)/2 \rfloor}$, for $i \in [n]$. Note that, $\deg(g_i) = k + \lfloor (i-1)/2 \rfloor \leq k + \lfloor (n-1)/2 \rfloor = n-1$. Define $n \times n$ matrix M_n such that

$$M_n \cdot [x]_n^T \;:=\; \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{bmatrix} .$$

It is easy to see that $g_1, g_3, g_5, \ldots$ are linearly independent over $\mathbb{R}$. Therefore, $\text{rank}(M_n) = \text{rank}_{\mathbb{R}}(g_1(x), \ldots, g_n(x)) = \lfloor (n-1)/2 \rfloor + 1 = \lfloor (n+1)/2 \rfloor$.

Define $A_n := M_n^T \cdot M_n$. By definition, A_n is PSD and $\text{rank}(A_n) = \lfloor (n+1)/2 \rfloor$. This follows from the classical fact that for any matrix A over $\mathbb{R}$, $\text{rank}(A^T A) = \text{rank}(A)$. Also A_n is *explicit* (entries are P-computable from definition). Now, assume there is some $m \times n$ matrix B such that $A_n = B^T \cdot B$. Then, denote $B[x]_n := \begin{bmatrix} \ell_1 & \ell_2 & \ldots & \ell_m \end{bmatrix}^T$, where $\ell_i \in \mathbb{R}[x]$ are univariate polynomials of degree at most $n-1$. Observe that number of *non-zero* entries in B is precisely $\sum_{i \in [m]} |\ell_i|_1$. Thus, it suffices to show that $\sum_{i \in [m]} |\ell_i|_1 \geq \Omega(n^2)$.

As $\text{rank}(B) = \text{rank}(B^T B) = \text{rank}(A_n) = \lfloor (n+1)/2 \rfloor$, we must have $m \geq \lfloor (n+1)/2 \rfloor$. Thus,

$$\begin{aligned} A_n \;=\; B^T \cdot B \;&\Longrightarrow\; [x]_n M_n^T \cdot M_n [x]_n^T \;=\; [x]_n B^T \cdot B [x]_n^T \\ &\Longleftrightarrow\; \sum_{i=1}^{n} g_i(x)^2 \;=\; \sum_{i=1}^{m} \ell_i^2 \\ &\Longleftrightarrow\; (x+1)^{2k} \cdot f(x) \;=\; \sum_{i=1}^{m} \ell_i^2 \qquad , \text{ where } f(x) \;:=\; \sum_{i=1}^{n} x^{2 \cdot \lfloor (i-1)/2 \rfloor} \\ &\Longrightarrow\; \sum_{i=1}^{m} |\ell_i|_1 \;\geq\; (\lfloor (n+1)/2 \rfloor) \cdot (k+1) \;\geq\; \frac{n^2}{4} \qquad \text{by Lemma 32.} \end{aligned}$$

$\square$

C.2 Lower bound for invertible circuits over $\mathbb{R}$: Proof of the second part of Thm.20

This subsection is devoted to proving the *second* part of Theorem 20. This proof uses SOS lower bound for a bivariate polynomial. Before that, we state a weak form of a classical lemma due to Descartes which will be used later.

Lemma 34 (Descartes rule of signs)**.** *Let $p(x) \in \mathbb{R}[x]$ be a polynomial with t many monomials. Then, number of distinct positive roots in $p(x)$ can be at most $t-1$.*

Investigating sum of product of two polynomials is similar to looking at the SOS; as, one can write $f \cdot g = ((f+g)/2)^2 - ((f-g)/2)^2$. The summand fan-in at most doubles. Thus, proving lower bound for sum of product of two polynomials is 'same' as proving SOS lower bound. The following lemma proves certain lower bound on sum of sparsity when a specific *bivariate* polynomial is written as sum of product of two polynomials (with certain restrictions).

Lemma 35. *Let $f_d := f_{d,t}(x,y) := \left(\prod_{i\in[d]}(x-i)(y-i)\right) \cdot p(x,y)$, for some polynomial $p \in \mathbb{R}[x,y]$ such that $deg_x(p) = deg_y(p) = t$. Suppose, $f_d = \sum_{i\in[d+t+1]} \ell_i(x) \cdot \tilde{\ell}_i(y)$, where $\ell_i, \tilde{\ell}_i$'s are polynomials of degree at most $d+t$; with the additional property that $\tilde{\ell}_1, \ldots, \tilde{\ell}_{d+t+1}$ are $\mathbb{R}$-linearly independent.*

Then, $\sum_{i=1}^{d+t+1} |\ell_i|_1 \;\geq\; m \cdot (d+1)$, where m is the number of non-zero ℓ_i.

Proof. Suppose, $g(x) := \gcd(\ell_1, \ldots, \ell_{d+t+1})$. We claim that $\prod_{i=1}^{d}(x-i) \mid g(x)$. Note that, it suffices to prove the claim; as, $\prod_{i=1}^{d}(x-i) \mid \ell_i(x)$ for each non-zero ℓ_i implies $|\ell_i|_1 \geq d+1$ by Lemma 34.

We prove the claim by contradiction. Suppose, there exists $j \in [d]$ such that $x - j \nmid g(x)$, so $g(j) \neq 0$. Fix this j. Hence, there exists i such that $\ell_i(j) \neq 0$.

In particular, $\boldsymbol{v} := \begin{bmatrix}\ell_1(j) & \ell_2(j) & \ldots & \ell_{d+t+1}(j)\end{bmatrix}^T \neq \mathbf{0}$. Define the $(d+t+1) \times (d+t+1)$ matrix A as

$$[y]_{d+t+1} \cdot A \;:=\; \begin{bmatrix}\tilde{\ell}_1 & \tilde{\ell}_2 & \ldots & \tilde{\ell}_{d+t+1}\end{bmatrix}\text{ , where } [y]_{d+t+1} \;:=\; \begin{bmatrix}1 & y & \ldots & y^{d+t}\end{bmatrix}\;.$$

Observe: $\mathrm{rank}_{\mathbb{R}}(\tilde{\ell}_1, \ldots, \tilde{\ell}_{d+t+1}) = d+t+1 \iff A$ is invertible. But,

$$\begin{aligned}
\boldsymbol{v} \neq \mathbf{0} \text{ and } A \text{ is invertible } &\implies A \cdot \boldsymbol{v} \;\neq\; \mathbf{0} \\
&\implies [y]_{d+t+1} \cdot A\boldsymbol{v} \;\neq\; 0 \\
&\implies \sum_{i=1}^{d+t+1} \tilde{\ell}_i(y) \cdot \ell_i(j) \;\neq\; 0 \\
&\implies f_{d,t}(j,y) \;\neq\; 0
\end{aligned}$$

which is a contradiction! Therefore, $\prod_{i=1}^{d}(x-i) \mid g(x)$ and so we are done. □

Recall that an *invertible* depth-2 circuit computes a matrix A such that whenever $A = BC$, either B or C has to be invertible. We prove the *second* part of Theorem 20.

Theorem 36 (Reproving Thm.1.5 of [KV19])**.** *There exists an explicit family of $n \times n$ PSD matrices $\{A_n\}_{n\in\mathbb{N}}$ such that every invertible circuit over $\mathbb{R}$ computing A_n has size $\Omega(n^2)$.*

Proof. Denote $k := \lfloor n/2 \rfloor$. Define $g_i(x) := \prod_{i=1}^{k}(x-i) \cdot x^{\lfloor (i-1)/2 \rfloor}$, for $i \in [n]$. Note that $\deg(g_i) = k + \lfloor (i-1)/2 \rfloor \leq k + \lfloor (n-1)/2 \rfloor = n-1$. Define the $n \times n$ matrix M_n as

$$M_n \cdot [x]_n^T \;:=\; \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_n(x) \end{bmatrix}\;.$$

It is easy to see that $g_1, g_3, g_5, \ldots$ are linearly independent over $\mathbb{R}$. Therefore, $\mathrm{rank}(M_n) = \mathrm{rank}_{\mathbb{R}}\,(g_1(x), \ldots, g_n(x)) = \lfloor (n-1)/2 \rfloor + 1 = \lfloor (n+1)/2 \rfloor$.

Define $A_n := M_n^T \cdot M_n$. By definition, A_n is PSD and $\mathrm{rank}(A_n) = \lfloor (n+1)/2 \rfloor$. This follows from the classical fact that for any matrix A, $\mathrm{rank}(A^T A) = \mathrm{rank}(A)$ over $\mathbb{R}$. Also A_n is *explicit* (entries are P-computable from definition).

Suppose, there exists $n \times n$ invertible matrix B and some $n \times n$ matrix C such that $A_n = B \cdot C$ (the other case where C is invertible is similar). Note that, from classical property of rank of matrices, $\mathrm{rank}(C) \geq \mathrm{rank}(A_n) = \lfloor (n+1)/2 \rfloor$. With the usual notation of $[x]_n$ and $[y]_n$ used before, denote

$$[y]_n \cdot B := \begin{bmatrix} \tilde{\ell}_1(y) & \tilde{\ell}_2(y) & \ldots & \tilde{\ell}_n(y) \end{bmatrix} \text{ and } C \cdot [x]_n^T := \begin{bmatrix} \ell_1(x) & \ell_2(x) & \ldots & \ell_n(x) \end{bmatrix}^T .$$

Note that the degree of each $\ell_i, \tilde{\ell}_i$ can be at most $n-1$. Thus,

$$\begin{aligned} A_n = B \cdot C \implies [y]_n M_n^T \cdot M_n [x]_n^T &= [y]_n \cdot B \cdot C \cdot [x]_n^T \\ \iff \sum_{i=1}^{n} g_i(x) \cdot g_i(y) &= \sum_{i=1}^{n} \ell_i(x) \cdot \tilde{\ell}_i(y) \\ \iff \left(\prod_{i=1}^{k} (x-i)(y-i) \right) \cdot p(x,y) &= \sum_{i=1}^{n} \ell_i(x) \cdot \tilde{\ell}_i(y) \end{aligned}$$

where $p(x,y) := \sum_{i \in [n]} (xy)^{\lfloor (i-1)/2 \rfloor}$. The LHS is actually of the form $f_{k,\lfloor (n-1)/2 \rfloor}(x,y)$ as in Lemma 35. From the lower bound on rank of C, we know that there must be at least $\lfloor (n+1)/2 \rfloor$ many non-zero ℓ_i's. Therefore, Lemma 35 gives us $\sum_{i=1}^{n} |\ell_i|_1 \geq \lfloor (n+1)/2 \rfloor \cdot (k+1) \geq n^2/4$. □

C.3 Newton polygon and bivariate SOS lower bound

Consider a bivariate polynomial $f \in \mathbb{F}[X,Y]$. To each monomial $X^i Y^j$ appearing in f with a nonzero coefficient, we associate a point with coordinate (i,j) in the Euclidean plane. Let $\mathrm{Mon}(f)$ denotes this finite set of points. If A is a set of points in the plane, we denote by $\mathrm{conv}(A)$ the *convex hull* of A. By definition, the *Newton polygon* of f, denoted by $\mathrm{Newt}(f)$, is the convex hull of $\mathrm{Mon}(f)$, i.e., $\mathrm{Newt}(f) = \mathrm{conv}(\mathrm{Mon}(f))$. Note that $\mathrm{Newt}(f)$ has at most t edges if f has t monomials. The following result is well known in the literature.

Theorem 37. [Ost75] *$Newt(fg) = Newt(f) + Newt(g) := \{p + q \mid p \in Newt(f), q \in Newt(g)\}$.*

From the above theorem, one can deduce that $\mathrm{Newt}(f^2) = 2 \cdot \mathrm{Newt}(f)$. But, if S is a convexly independent subset of $2 \cdot \mathrm{Newt}(f)$, how large can S be? [A set is called *convexly independent* if its elements are exactly the vertices of its convex hull.]

This will be crucial in the next section. Here is an important theorem (which is optimal up to constant factors) regarding the size of S; compare the bound with the trivial mn.

Theorem 38. [EPRS08] *Let P and Q be two planar point sets with $|P| = m$ and $|Q| = n$. Let S be a convexly independent subset of the Minkowski sum $P + Q$. Then, we have $|S| \leq O(m^{2/3} n^{2/3} + m + n)$.*

Corollary 39. *Let P be a planar point set with $|P| = n$. Let S be a convexly independent subset of rP (r is a constant). Then, $|S| \leq O(n^{r^{\log(4/3)}})$.*

Proof. Let $T(r)$ be the maximum size of convexly independent subset of rP. Thus, we must have $T(r) \leq O(T(r/2)^{4/3})$ with $T(1) \leq n$. Thus, $T(r) \leq O(n^{(4/3)^{\log r}}) = O(n^{r^{\log(4/3)}})$. □

Using convexity theory, we establish the lower bound of $\Omega(d^{1/r^{\log(4/3)}})$ for the bivariate polynomial $\sum_{i=0}^{d} x^i y^{i^2}$. This polynomial was studied in [KPTT15].

Theorem 40. *For $f(x,y) := \sum_{i=0}^{d} x^i y^{i^2}$, we have $S_{\mathbb{R}}(f,r,s) \geq \Omega(d^{1/r^{\log(4/3)}})$, for any $s \geq 1$ and constant r.*

Proof sketch. Write $f(x,y) = \sum_{i\in[s]} \ell_i(x,y)^r$. Let S_i be the set of points in the plane corresponding to the monomials of ℓ_i^r which appear in f with a nonzero coefficient. Since $\mathrm{Newt}(f)$ is the convex hull of $\cup_i \mathrm{conv}(S_i)$, it is enough to bound the number of vertices of $\mathrm{conv}(S_i)$.

Of course, the vertices of $\mathrm{conv}(S_i)$ is a convexly independent subset of $\mathrm{Mon}(\ell_i^r) \subseteq r\mathrm{Mon}(\ell_i)$. Hence, by Corollary 39, we get that $\mathrm{conv}(S_i)$ has at most $O(|\ell_i|_1)^{r^{\log(4/3)}}$ many vertices. Thus, the convex hull of $\bigcup_i \mathrm{conv}(S_i)$ has at most $O(\sum_i ||\ell_i||^{r^{\log(4/3)}})$ vertices. On the other hand, as $y = x^2$ is a convex function, $\mathrm{Newt}(f)$ has $d+1$ many vertices. Therefore,

$$\sum_i (|\ell_i|_1)^{r^{\log(4/3)}} \;\geq\; d+1 \;\implies\; \sum_i |\ell_i|_1 \;\geq\; \Omega(d^{1/r^{\log(4/3)}}) \;.$$

By definition, we must have $S_{\mathbb{R}}(f,r,s) \geq \Omega(d^{1/r^{\log(4/3)}})$, for any $s \geq 1$. $\square$

Remark. As $\log(4/3) \approx 0.415 < 1$, the above is a better lower bound on $S_{\mathbb{R}}(\cdot)$ than the trivial lower bound of $\mathrm{sparsity}(f)^{1/r} = (d+1)^{1/r}$.

D Large sparsity-sum measure also implies Theorems 2–4

Recall the measure $S_{\mathbb{F}}(\cdot)$ defined in Section 3.5. By definition, $U_{\mathbb{F}}(\cdot) \leq S_{\mathbb{F}}(\cdot)$. Let field $\mathbb{F}$ be characteristic zero, or a finite field of characteristic $> r$. Thus, one can, as well conjecture the following.

Conjecture 3 (C3)**.** *There exist positive constants $\delta_1 \leq 1$, $\delta_2 \geq 1$ and a constant prime-power r such that $S_{\mathbb{F}}\left(f_d, r, d^{\delta_1}\right) \geq d/r^{\delta_2}$, for all large enough $d \in I_r$.*

$S_{\mathbb{F}}$ is large for *'random'* polynomials f. We can easily argue that for a *random* polynomial f and for some $s \geq 1$, if $f = \sum_{i\in[s]} c_i \ell_i^r$, then $\sum_{i\in[s]} |\ell_i|_1 \geq \Omega(|f|_1)$ (implying $S_{\mathbb{F}}$ large). One can view $\ell_i \in \mathbb{F}[z_{i1}, \ldots, z_{it_i}][\boldsymbol{x}]$, where the transcendence degree (*tr.deg*) of the coefficient polynomials of ℓ_i is t_i. Observe that, this means $|\ell_i|_1 \geq t_i$, for all i. As coefficient of ℓ_i^r is generated by the coefficient of ℓ_i, the tr.deg of coefficient polynomials produced in RHS is at most $\sum_i (t_i + 1)$. Now, as f is a 'random' polynomial, the coefficients of f are algebraically independent. In particular, the tr.deg of the coefficient polynomials of f is $|f|_1$. Thus, $\sum_{i\in[s]}(|\ell_i|_1 + 1) \geq \sum_i (t_i + 1) \geq |f|_1$; implying that $S_{\mathbb{F}}(f,r,s) \geq \Omega(|f|_1)$.

One can show that Theorems 2–4 hold true assuming conjecture C3 (*instead* of conjecture C1). The analogous proof for Theorem 2 (resp. Thm.4) is identical to the original one. However, the proof of Theorem 3 slightly differs. For the sake of completeness, we give a proof sketch.

Theorem 41. *If GRH and Conjecture C3, for some $r \geq 25$, hold, then* VP $\neq$ VNP*.*

Proof sketch. Let Conjecture C3 be true for a fixed $r \geq 25, \delta_1, \delta_2$. Define $\delta_1' \leq \min(0.71, \delta_1)$.We consider *non-constant* n and let $\boldsymbol{x} := (x_1, \ldots, x_n)$. For all large $n \in \mathbb{N}$, there exists *exactly* one $d := d(n)$ such that $d \in I_r \cap [(2^n - 1)/(r+1), 2^n - 1]$. Thus, $n = \Theta(\log d)$.

Define the polynomial family $P_n(\boldsymbol{x}) := \psi_{n,d}(f_d)$, via the inverse Kronecker map applied on $f_d = (x+1)^d$. Note that, P_n is an n-variate *multilinear* polynomial. This is ensured because the individual degree $d_n := \lceil (d+1)^{1/n} \rceil - 1 \leq 1$ ($\because d \leq 2^n - 1$).

First, we claim: if GRH holds and VP = VNP, then $\{P_n\}_n \in$ VP.

Conditionally, $\{\mathbf{P_n}\}_{\mathbf{n}} \in$ VP : This part of the proof is same as proof of Theorem 3.

Next, we show: Conjecture C3 implies that, for $\mu > 3r/\delta_1'$, $\mathrm{size}(P_n) > d^{1/\mu} = 2^{\Omega(n)}$. Thus, $\{P_n\}_n \notin$ VP. This contradiction would finish the proof of Theorem 41.

Conditionally, $\{\mathbf{P_n}\}_\mathbf{n} \notin \mathsf{VP}$: This proof is very similar to the hardness part of Theorem 3. Except the parameter setting is slightly different ($\because$ sparsity-sum is usually larger than support-union), so we need to go through the details. We prove that $\text{size}(P_n) > d^{1/\mu} = 2^{\Omega(n)}$, where $\mu > 3r/\delta_1'$. Assume that this is not the case, then there exists an infinite subset $J \subset \mathbb{N}$ such that the algebraic circuit complexity of $\{P_n(\boldsymbol{x})\}_n$ is $\leq d^{1/\mu}$, for $n \in J$. Consider $J_r := \{d(n) \in I_r \mid n \in J\} \subseteq I_r$.

We use similar depth-4 reduction (with top/ bottom $\Sigma\Pi$ parts analysis as in the proof). Fix the t such that $5^t \leq r < 5^{t+1}$. We know that P_n can be written as $P_n = \sum_{i=1}^{\tilde{s}} \tilde{c}_i \cdot \tilde{g}_i^r$, where $\tilde{s} := s_0 \cdot 2^{5^t} \cdot (r+1)$ with $s_0 = \binom{s'+5^t}{5^t}$ where $s' \leq d^{3/\mu} \cdot n^6$ and each $\tilde{g}_i$ is an n-variate polynomial of degree $\leq n/2^t$. So, we bound the sparsity-sum $\sum_{i \in [\tilde{s}]} |\tilde{g}_i|_1 \leq \tilde{s} \cdot s_1$ where $s_1 := \binom{n+n \cdot 2^{-t}}{n}$. Apply the Kronecker map $\phi_{n,1}$ on P_n. As, $\phi_{n,1} \circ \psi_{n,d} = \text{id}$, over $\mathbb{F}[x]^{\leq d}$, we get

$$f_d \;=\; (\phi_{n,1} \circ \psi_{n,d})(f_d) \;=\; \phi_{n,1}(P_n) \;=\; \sum_{i=1}^{\tilde{s}} \tilde{c}_i \cdot \left(\phi_{n,1}\left(\tilde{g}_i\right)\right)^r \;.$$

As Kronecker substitution can not increase the sparsity, we have $\sum_{i \in [\tilde{s}]} |\phi_{n,1}\left(\tilde{g}_i\right)|_1 \leq \tilde{s} \cdot s_1$. Thus, we have $S_{\mathbb{F}}(f_d, r, \tilde{s}) \leq \tilde{s} \cdot s_1$, from the definition of $S_{\mathbb{F}}(\cdot)$. Using details of the proof of Theorem 3, we know that $\tilde{s} < d^{\delta_1'}$, when $\mu > 3r/\delta_1'$; and $s_1 < 13.6^{n/4}$, for all large enough n. We claim that $\tilde{s} \cdot s_1 < d/r^{\delta_2}$, for large d. Then, we shall have $S_{\mathbb{F}}(f_d, r, d^{\delta_1}) < d/r^{\delta_2}$ over an infinite subset $J_r \subseteq I_r$; which obviously contradicts Conjecture C1.

To prove the bound on $\tilde{s} \cdot s_1$, note that

$$\tilde{s} \cdot s_1 \;<\; d^{\delta_1'} \cdot (13.6)^{n/4} \;<\; 2^{n \cdot (\delta_1' + \log(13.6)/4)} \;<\; o(d) \;.$$

We used the fact: $\log(13.6)/4 < 0.284$, $\delta_1' \leq 0.71$, and $2^n = \Theta(d)$. In particular, we deduce: $\tilde{s} \cdot s_1 < d/r^{\delta_2}$, for all large enough d.

This gives us: $S_{\mathbb{F}}(f_d, r, d^{\delta_1}) < d/r^{\delta_2}$, over an infinite subset $J_r \subseteq I_r$. The contradiction therefore implies: algebraic circuit complexity of $\{P_n(\boldsymbol{x})\}_n$ is $> d^{1/\mu} = 2^{\Omega(n)}$. So, $\{P_n\}_n \notin \mathsf{VP}$.

Eventually, the above two conclusions about $\{P_n\}_n$ are contradictory; thus, we get $\mathsf{VP} \neq \mathsf{VNP}$ under GRH and Conjecture C3. $\square$

Remark. We are not able to make the proof of Theorem 1 work with sparsity-sum measure. Our naive attempt, to construct a k (=constant) variate P_n from f_d, does not give the hardness required to design an HSG.