Electronic Colloquium on Computational Complexity, Report No. 99 (2024)

A subquadratic upper bound on Hurwitz's problem and related non-commutative polynomials*

Pavel Hrubeš †

June 5, 2024

Abstract

For every n, we construct a sum-of-squares identity

$$(\sum_{i=1}^{n} x_i^2)(\sum_{j=1}^{n} y_j^2) = \sum_{k=1}^{s} f_k^2 \,,$$

where f_k are bilinear forms with complex coefficients and $s = O(n^{1.62})$. Previously, such a construction was known with $s = O(n^2/\log n)$. The same bound holds over any field of positive characteristic.

As an application to complexity of non-commutative computation, we show that the polynomial $\mathsf{ID}_n = \sum_{i,j\in[n]} x_i y_j x_i y_j$ in $2n$ non-commuting variables can be computed by a non-commutative arithmetic circuit of size $O(n^{1.96})$. This holds over any field of characteristic different from two. The same bound applies to non-commutative versions of the elementary symmetric polynomial of degree four and the rectangular permanent of a $4 \times n$ matrix.

1 Introduction

The problem of Hurwitz [14] asks for which integers n, m, s does there exist a sum-of-squares identity

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_m^2) = f_1^2 + \cdots + f_s^2 \,, \tag{1}$$

where $f_1, \ldots, f_s$ are bilinear forms in x and y with complex coefficients. Historically, the problem was motivated by existence of non-trivial identities with $n = m = s$. Starting with the obvious $x_1^2 y_1^2 = (x_1 y_1)^2$, the first remarkable identity is

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2 \,.$$

*The first set of results has appeared in [10] and was accepted to CCC'24.

†Institute of Mathematics of ASCR, pahrubes@gmail.com. This work was supported by Czech Science Foundation GAČR grant 19-27871X.

ISSN 1433-8092

It can be interpreted as asserting multiplicativity of the norm on complex numbers. Euler's 4-square identity is an example with $n, m, s = 4$ which has later been interpreted as multiplicativity of the norm on quaternions. The final one is an 8-square identity which arises in connection to the algebra of octonions.

A classical result of Hurwitz [14] shows that these are the only cases: an identity (1) exists with $m, s = n$ iff $n \in \{1, 2, 4, 8\}$. An extension of this result is given by Hurwitz-Radon theorem [18]: an identity (1) exists with $s = n$ iff $m \leq \rho(n)$, where $\rho(n)$ is the Hurwitz-Radon number. The value of $\rho(n)$ is known exactly. For every n, $\rho(n) \leq n$ and equality is achieved only in the cases $n \in \{1, 2, 4, 8\}$. Asymptotically, $\rho(n)$ lies between $2 \log_2 n$ and $2 \log_2 n + 2$ if n is a power of 2. As shown in [19], Hurwitz-Radon theorem remains valid over any field of characteristic different from two. Hurwitz's problem is an intriguing question with connections to several branches of mathematics. We recommend D. Shapiro's monograph [20] on this subject.

Let $\sigma(n)$ denote the smallest s such that an identity (1) with $m = n$ exists. While Hurwitz-Radon theorem solves the case $s = n$ *exactly*, even the *asymptotic* behavior of $\sigma(n)$ is not known. Elementary bounds[1] are $n \leq \sigma(n) \leq n^2$. Hurwitz's theorem implies that the first inequality is strict if n is sufficiently large. Using Hurwitz-Radon theorem, the upper bound can be improved to

$$\sigma(n) \leq O(n^2/\log n)\,.$$

As far as we are aware, this was the best asymptotic upper bound previously known. In this paper, we will improve it to a truly subquadratic bound

$$\sigma(n) \leq O(n^{1.62})\,. \tag{2}$$

A specific motivation for this problem comes from arithmetic circuit complexity. In [11], Wigderson, Yehudayoff and the current author related the sum-of-squares problem with the complexity of non-commutative computations. Non-commutative arithmetic circuit is a model for computing polynomials whose variables do not multiplicatively commute. Since the seminal paper of Nisan [17], it has been an open problem to give a superpolynomial lower bound on circuit size in this model. In [11], it has been shown that a superlinear lower bound on $\sigma(n)$ of the form $\Omega(n^{1+\epsilon})$, $\epsilon > 0$, translates to an exponential circuit lower bound in the non-commutative setting. More specifically, such a lower bound on σ implies an $\Omega(n^{1+\epsilon})$ lower bound for the degree four polynomial

$$\mathsf{ID}_n = \sum_{i,j\in[n]} x_i y_j x_i y_j\,,$$

which in turn can be lifted to an exponential lower bound for an explicit polynomial of degree n. Hence, providing asymptotic lower bounds on Hurwitz's problem can be seen as a concrete approach towards answering Nisan's question. A more general, and hence less concrete, result of this flavor was given

[1] The former is obtained by substituting $(1, 0, \ldots, 0)$ for the y variables, the latter by writing $(\sum x_i^2)(\sum_j y_j^2) = \sum_{i,j}(x_i y_j)^2$.

by Carmosino et al. in [4]. In an attempt to implement the sum-of-squares approach, the authors from [11] also gave an $\Omega(n^{6/5})$ lower bound under the assumption that the identity (1) involves *integer* coefficients only [12].

In view of previously known bounds on σ, it was conceivable that ID_n requires non-commutative arithmetic circuit of size $n^{2-o(1)}$. However, we will use the upper bound (2) to construct a circuit for ID_n of a subquadratic size. The same applies to related polynomials such as the non-commutative elementary symmetric polynomial $S_{4,n}$ or the rectangular permanent of a $4\times n$ matrix. The latter polynomials have been previously studied by Arvind et al. [1], see also [22]. The circuit bound we obtain for ID_n is quantitatively weaker than (2). This is partly because the construction uses matrix multiplication as an ingredient. To determine the complexity of matrix multiplication is a fundamental open problem in its own right. We will use bounds on rectangular matrix multiplication provided by le Gall and Urrutia [5] where this exciting problem is discussed further.

The upper bounds presented here go against the lower bound approach of [11]. Since the bounds are superlinear, they do not immediately frustrate the approach, but rather dampen its optimism.

2 Main results

Let $\mathbb{F}$ be a field. Define $\sigma_{\mathbb{F}}(n,m)$ as the smallest s such that there exist bilienear[2] $f_1,\dots,f_s \in \mathbb{F}[x_1,\dots,x_n,y_1,\dots y_m]$ satisfying (1). Furthermore, let $\sigma_{\mathbb{F}}(n) := \sigma_{\mathbb{F}}(n,n)$.

Theorem 1. *Let $\mathbb{F}$ be a field containing a square root of -1 or a field of positive characteristic. Then $\sigma_{\mathbb{F}}(n) \leq O(n^c)$ where $c < 1.62$.*

This will be proved in Section 4. This implies that over *any field*, we can write (see Section 5.1)

$$(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2) = f_1^2 + \cdots + f_s^2 - (f_{s+1}^2 + \cdots + f_{2s}^2)\,,$$

with $s \leq O(n^c)$ and $f_1,\dots,f_{2s}$ bilinear.

Remark 2. *If the field has characteristic two, Theorem 1 is trivial. Since $(\sum_i x_i^2)(\sum_j y_j^2) = (\sum_{i,j} x_i y_j)^2$, we have $\sigma_{\mathbb{F}}(n,m) = 1$.*

We will give an application to complexity of non-commutative polynomials. A *non-commutative polynomial* over $\mathbb{F}$ is a formal sum of products of variables and field elements. We assume that the variables do not multiplicatively commute, whereas they commute additively, and with elements of $\mathbb{F}$. A *non-commutative arithmetic circuit* is a standard model for computing such

[2] Namely, of the form $\sum_{i,j} a_{i,j} x_i y_j$.

polynomials. A non-commutative circuit ψ can be defined as a directed acyclic graph as follows. Nodes (or gates) of in-degree zero are labelled by either a variable or a field element in $\mathbb{F}$. All the other nodes have in-degree two and they are labelled by either $+$ or $\times$. The two edges going into a gate labelled by $\times$ are labelled by *left* and *right* to indicate the order of multiplication. Every node in ψ computes a non-commutative polynomial in the obvious way. We say that ψ computes a polynomial f if there is a gate in ψ computing f. As the size of ψ, we take the number of its vertices.

The *identity polynomial* is a polynomial in $2n$ non-commuting variables

$$\mathsf{ID}_n = \sum_{i,j\in[n]} x_i y_j x_i y_j \,.$$

It can trivially be computed by a non-commutative circuit of a quadratic size. We also consider non-commutative versions of the elementary symmetric polynomial $S_{k,n}$ and the rectangular permanent of a $k\times n$ matrix

$$S_{k,n} = \sum_{(i_1,\dots,i_k)} x_{i_1}\cdots x_{i_k}\,, \quad \mathsf{perm}_{k,n} = \sum_{(i_1,\dots,i_k)} x_{1,i_1}\cdots x_{k,i_k}\,,$$

where the sums range over ordered k-tuples $(i_1,\dots,i_k)$ where $i_1,\dots,i_k$ are pairwise distinct elements of $[n]$.

Theorem 3. *Over a field of characteristic different from two, ID_n, $S_{4,n}$ and $\mathsf{perm}_{4,n}$ can be computed by a non-commutative circuit of size $O(n^c)$ where $c < 1.96$.*

Theorem 3 will be proved as Theorem 25 and Corollary 27 in Section 6.

Remark 4. *The division of variables in ID_n into two parts is a cosmetic detail intended to match the format of Hurwitz's problem. The non-commutative complexity of $\sum_{i,j} x_i x_j x_i x_j$, ID_n, and $\sum_{i,j} x_i y_j z_i u_j$ differ by a constant factor only (cf. [11]). What is crucial is the order of multiplication: both $\sum_{i,j} x_i x_i y_j y_j$ and $\sum_{i,j} x_i y_j y_j x_i$ have a non-commutative circuit of a linear size.*

Notation Given vectors $u,v\in\mathbb{F}^n$, $\langle u,v\rangle := \sum_{i=1}^n u_i v_i$ is their inner product. For a set S, $\binom{S}{k}$ denotes the set of k-element subsets of S and $\binom{S}{\le k}$ the set of subsets with at most k elements. $\binom{n}{\le k} := \sum_{i=0}^k \binom{n}{i}$. $[n]$ is the set $\{1,\dots,n\}$.

3 Hurwitz-Radon conditions

In this section, we give some well-known properties of σ that we will need later.

The definition immediately implies thet $\sigma_{\mathbb{F}}(n,m)$ is symmetric, subadditive, and monotone:

$$\begin{aligned} \sigma_{\mathbb{F}}(n,m) &= \sigma_{\mathbb{F}}(m,n)\,,\\ \sigma_{\mathbb{F}}(n,m_1+m_2) &\le \sigma_{\mathbb{F}}(n,m_1)+\sigma_{\mathbb{F}}(n,m_2)\,,\\ \sigma_{\mathbb{F}}(n,m) &\le \sigma_{\mathbb{F}}(n,m')\,,\ m\le m'\,. \end{aligned} \tag{3}$$

The following lemma gives a characterization of σ in terms of Hurwitz-Radon conditions (4). A proof can be found, e.g., in [20], but we present it for completeness.

Lemma 5. *Let $\mathbb{F}$ be a field of characteristic different from two. Then $\sigma_{\mathbb{F}}(n,m)$ equals the smallest s such that there exist matrices $H_1,\dots,H_m \in \mathbb{F}^{n\times s}$ satisfying*

$$\begin{aligned} &H_iH_i^t = I_n\,,\\ &H_iH_j^t + H_jH_i^t = 0\,,\ i\neq j\,, \end{aligned} \tag{4}$$

for every $i,j\in[m]$.

Proof. Let $f_1,\dots,f_s$ be bilinear polynomials in variables $x_1,\dots,x_n$ and $y_1,\dots,y_m$. Then the vector $\bar f = (f_1,\dots,f_s)$ can be written as

$$\bar f = \sum_{i=1}^{n} \bar x H_i y_i\,,$$

where $\bar x = (x_1,\dots,x_n)$ and $H_i\in\mathbb{F}^{n\times s}$. Hence

$$\sum_{k=1}^{s} f_k^2 = \bar f\bar f^t = \sum_i y_i^2 \bar x H_iH_i^t\bar x^t + \sum_{i<j} y_iy_j\bar x(H_iH_j^t + H_jH_i^t)\bar x^t\,.$$

If the matrices satisfy (4), this equals $\sum_i y_i^2\bar x I_n\bar x^t = (y_1^2+\cdots+y_m^2)(x_1^2+\cdots+x_n^2)$, which gives a sum-of-squares identity with s squares. Conversely, if $(y_1^2+\cdots+y_m^2)(x_1^2+\cdots+x_n^2) = \sum f_k^2$, we must have $\bar xH_iH_i^t\bar x^t = x_1^2+\cdots+x_n^2$ and $\bar x(H_iH_j^t+H_jH_i^t)\bar x^t = 0$. In characteristic different from 2, this is possible only if the conditions (4) are satisfied. $\square$

Given a natural number of the form $n = 2^k a$ where a is odd, the Hurwitz-Radon number is defined as

$$\rho(n) = \begin{cases} 2k+1\,, & \text{if } k=0\\ 2k\,, & \text{if } k=1\\ 2k\,, & \text{if } k=2\\ 2k+2\,, & \text{if } k=3 \end{cases} \ \mathsf{mod}\, 4$$

Observe that

$$2\log_2 n \le \rho(n) \le 2\log_2(n)+2\,,$$

whenever n is a power of two.

Square matrices A_1,A_2 *anticommute* if $A_1A_2 = -A_2A_1$. A family of square matrices $A_1,\dots,A_t$ will be called *anticommuting* if A_i, A_j anticommute for every $i\neq j$.

The following lemma is a key ingredient in the proof of Hurwitz-Radon theorem. A self-contained construction can be found in [6].

Lemma 6. *For every n, there exists an anticommuting family of $t = \rho(n) - 1$ integer matrices $e_1, \dots, e_t \in \mathbb{Z}^{n \times n}$ which are orthonormal and antisymmetric (i.e., $e_i e_i^t = I_n$ and $e_i = -e_i^t$).*

Remark 7. *A straightforward construction (see, e.g., [9]) gives an anticommuting family of $t = 2\log_2 n + 1$ integer matrices $e_1, \dots, e_t \in \mathbb{Z}^{n \times n}$ with $e_i^2 = \pm I_n$ whenever n is a power of two. With minor modifications, these matrices could be used in the subsequent construction instead.*

4 The construction

Let $e_1, \dots, e_t$ be a set of square matrices. Given $A = \{i_1, \dots, i_k\} \subseteq [t]$ with $i_1 < \dots < i_k$, let $e_A := \prod_{j=1}^{k} e_{i_j}$.

Lemma 8. *Let $e_1, \dots, e_t$ be a set of anticommuting matrices. If $A, B \subseteq [t]$ have even size (resp. odd size) then e_A, e_B anticommute assuming $|A \cap B|$ is odd (resp. even).*

Proof. Since e_i anticommutes with every e_j, $j \neq i$, but commutes with itself, we obtain

$$e_A e_i = (-1)^{|A \setminus \{i\}|} e_i e_A \,.$$

This implies that

$$e_A e_B = (-1)^q e_B e_A \,,$$

where $q = |A| \cdot |B| - |A \cap B|$. Hence if A, B are even (resp. odd) and their intersection is odd (resp. even), q is odd and e_A, e_B anticommute. □

Given integers $0 \leq k \leq t$, a *(k,t)-parity representation* of dimension s over a field $\mathbb{F}$ is a map $\xi : \binom{[t]}{k} \to \mathbb{F}^s$ such that for every $A, B \in \binom{[t]}{k}$

$$\begin{aligned} &\langle \xi(A), \xi(A) \rangle = 1 \,, \\ &\langle \xi(A), \xi(B) \rangle = 0 \,, \text{ if } A \neq B \text{ and } (|A \cap B| = k \bmod 2) \,. \end{aligned} \tag{5}$$

Lemma 9. *Let $0 \leq k \leq t$. Over $\mathbb{C}$, there exists a (k,t)-parity representation of dimension $\binom{t}{\leq \lfloor k/2 \rfloor}$.*

More generally, assume that $\mathbb{F}$ is a field of characteristic different from two containing a subfield $\mathbb{F}'$ such that every element of $\mathbb{F}'$ is a sum of r squares in $\mathbb{F}$. Then there exists a (k,t)-parity representation of dimension $r\binom{t}{\leq \lfloor k/2 \rfloor}$.

We will first prove the lemma over $\mathbb{C}$, the latter part will be shown in Section 4.1.

Proof of Lemma 9 over $\mathbb{C}$. Let $0 \leq k \leq t$ be given and $d := \lfloor k/2 \rfloor$.

For $a \in \{0,1\}^t$, let $|a|$ be the number of ones in a. Recall that a polynomial is multilinear, if every variable in it has individual degree at most one. We first observe:

Claim 10. *There exists a multilinear polynomial $f \in \mathbb{Q}[x_1, \dots, x_t]$ of degree at most d such that for every $a \in \{0,1\}^t$*

$$f(a) = \begin{cases} 1\,, & \text{if } |a| = k \\ 0\,, & \text{if } |a| < k \text{ and } (|a| = k \bmod 2)\,. \end{cases} \tag{6}$$

Proof of Claim. Consider the polynomial

$$g(x_1, \dots, x_t) := c \prod_{0 \le i < k,\, i = k \bmod 2} (\sum_{j=1}^{t} x_j - i)\,.$$

Then g has degree d and we can choose $c \in \mathbb{Q}$ so that g satisfies (6). Since we care about inputs from $\{0,1\}^t$, g can be rewritten as a multilinear polynomial f of degree at most d. □

Since f is multilinear, we can write it as

$$f(x_1, \dots, x_t) = \sum_{C \in \binom{[t]}{\le d}} \alpha_C \prod_{i \in C} x_i\,,$$

where α_C are rational coefficients. Identifying a subset A of $[t]$ with its characteristic vector in $\{0,1\}^t$, we have

$$f(A) = \sum_{C \subseteq A} \alpha_C\,.$$

Let $s := \binom{t}{\le d}$. Given $A \in \binom{[t]}{k}$, let $\xi(A) \in \mathbb{C}^s$ be the vector whose coordinates are indexed by subsets $C \in \binom{[t]}{\le d}$ such that

$$\xi(A)_C = \begin{cases} (\alpha_C)^{1/2}\,, & \text{if } C \subseteq A \\ 0\,, & \text{if } C \not\subseteq A\,. \end{cases}$$

This guarantees

$$\langle \xi(A), \xi(B) \rangle = \sum_C \xi(A)_C \xi(B)_C = \sum_{C \subseteq A \cap B} \alpha_C = f(A \cap B)\,.$$

Hence conditions (6) translate to the desired properties of the map ξ. □

Combining Lemma 8 and 9, we obtain the following bound on σ:

Theorem 11. *Let n be a non-negative integer. Let $0 \le k \le \rho(n) - 1$ and $m := \binom{\rho(n)-1}{k}$. Then*

$$\sigma_{\mathbb{C}}(n, m) \le n \cdot \binom{\rho(n) - 1}{\le \lfloor k/2 \rfloor}\,.$$

If $\mathbb{F}$ is as in the assumption of Lemma 9 then

$$\sigma_{\mathbb{F}}(n, m) \le rn \cdot \binom{\rho(n) - 1}{\le \lfloor k/2 \rfloor}\,.$$

Proof. Let n, k, m be as in the assumption. Let $e_1, \dots, e_t$ be the matrices from Lemma 6 with $t = \rho(n) - 1$. Let ξ be the (k,t)-parity representation given by Lemma 9. For $A \in \binom{[t]}{k}$, let

$$H_A := e_A \otimes \xi(A)\,,$$

where e_A is defined as in Lemma 8, $\xi(A)$ is viewed as a row vector, and $\otimes$ is the Kronecker (tensor) product.

Note that each H_A has dimension $n \times (ns)$ where s is the dimension of the parity representation, and there are $m = \binom{t}{k}$ such matrices H_A. By Lemma 5, it is sufficient to show that the system of matrices $H_A, A \in \binom{[t]}{k}$, satisfies Hurwitz-Radon conditions (4).

We have

$$H_A H_B^t = (e_A e_B^t) \otimes (\xi(A)\xi(B)^t) = \langle \xi(A), \xi(B) \rangle \cdot e_A e_B^t\,.$$

Since every e_i is orthonormal, we have $e_A e_A^t = I_n$. (5) gives $\langle \xi(A), \xi(A) \rangle = 1$ and hence

$$H_A H_A^t = I_n\,.$$

If $A \neq B$ then

$$H_A H_B^t + H_B H_A^t = \langle \xi(A), \xi(B) \rangle \cdot (e_A e_B^t + e_B e_A^t)\,. \tag{7}$$

If $|A \cap B| = k \,\mathsf{mod}\, 2$ then $\langle \xi(A), \xi(B) \rangle = 0$ by (5) and hence (7) equals zero. If $|A \cap B| \neq k \,\mathsf{mod}\, 2$ then $e_A e_B^t + e_B e_A^t = 0$. This is because $e_A e_B = -e_B e_A$ by Lemma 8 and that, since e_i are antisymmetric, e_A, e_B are either both symmetric or both antisymmetric. Therefore (7) equals zero for every $A \neq B \in \binom{[t]}{k}$. $\square$

Remark 12. *(i). If -1 is a sum of r squares over $\mathbb{F}$ then every element of $\mathbb{F}$ is a sum of $r+1$ squares. This follows by noting $a = (\frac{a+1}{2})^2 - (\frac{a-1}{2})^2$. Hence if $\mathbb{F}$ contains a square root of -1, as in the case of Gaussian rationals $\mathbb{Q}(\mathrm{i})$, every element of $\mathbb{F}$ is a sum of 2 squares.*

(ii). It follows from Lagrange's four-square theorem that every element of $\mathbb{F}_p$ is a sum of four squares. Furthermore, every element of $\mathbb{F}_p$ has a square root in $\mathbb{F}_{p^2}$

Theorem 1 is an application of Theorem 11.

Proof of Theorem 1. Let $\mathbb{F}$ be field containing a square root of -1 or a field of a positive characteristic p. If $p = 2$, the statement of the theorem is trivial. Otherwise, due to Remark 12, we can apply Theorem 11 with $r = 4$.

Assume first that n is a power of 16. This gives $\rho(n) = 2\log_2(n) + 1$. Let k be the smallest integer with $n \leq \binom{2\log_2 n}{k} =: m$. From the previous theorem and monotonicity of σ (cf. (3)), we obtain

$$\sigma_{\mathbb{F}}(n) \leq \sigma_{\mathbb{F}}(n, m) \leq 4ns\,,$$

where $s := \binom{2\log_2 n}{\leq \lfloor k/2 \rfloor}$.

We have $k = 2(\alpha + \epsilon_n)\log_2 n$ where $\alpha \in (0, \frac{1}{2})$ is such that $H(\alpha) = 1/2$ (H is the binary entropy function) and $\epsilon_n \to 0$ as n approaches infinity. We also have

$$s \leq 2^{2H(\frac{\alpha+\epsilon_n}{2})\log_2 n} = n^{2H(\frac{\alpha}{2})+\epsilon'_n},$$

where $\epsilon'_n \to 0$. Hence

$$\sigma_{\mathbb{F}}(n) \leq 4n^{1+2H(\frac{\alpha}{2})+\epsilon'_n}.$$

The numerical value of α is $0.11\ldots$ which leads to $\sigma_{\mathbb{F}}(n) \leq 4n^{1.615+\epsilon'_n} \leq O(n^{1.616})$.

If n is not a power of 16, take n' with $n < n' < 16n$ which is. By monotonicity of σ, we have $\sigma_{\mathbb{F}}(n) \leq \sigma_{\mathbb{F}}(n')$. □

4.1 The general case of Lemma 9

We now prove the remaining case of Lemma 9. The first objective is to reprove Claim 10 in positive characteristic.

Given non-negative integers $\bar{n} = (n_1, \ldots, n_d)$ let $B(\bar{n})$ be the $d \times d$ matrix $\{B(\bar{n})_{i,j}\}_{i,j\in[d]}$ with

$$B(\bar{n})_{i,j} = \binom{n_j}{i-1}.$$

We assume that $\binom{n}{k} = 0$ whenever $n < k$; this guarantees $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$.

Lemma 13. *If $\bar{n} = (r, r+2, \ldots, r+2(d-1))$ for some non-negative integer r then* $\det(B(\bar{n})) = 2^{\binom{d}{2}}$.

Proof. We claim that

$$\det(B(\bar{n})) = (\prod_{i=1}^{d-1} i!)^{-1} \det(V(\bar{n})),$$

where $V(\bar{n})$ is the Vandermonde matrix with entries $V(\bar{n})_{i,j} = n_j^{i-1}$. To see this, multiply every i-th row of $B(\bar{n})$ by $(i-1)!$ to obtain matrix $B'(\bar{n})$ with

$$\det(B'(\bar{n})) = (\prod_{i=1}^{d-1} i!) \det(B(\bar{n})).$$

An i-th row r_i of $B'(\bar{n})$ is of the form $(n_1^{i-1} + g_i(n_1), \ldots, n_d^i + g_i(n_d))$ where g_i is a polynomial of degree $< (i-1)$. This means that r_i equals the i-th row of $V(\bar{n})$ plus a suitable linear of combination of the preceding rows of $V(\bar{n})$. Therefore, $\det(B'(\bar{n})) = \det(V(\bar{n}))$.

Given $\bar{n}$ as in the assumption, we obtain

$$\det(V(\bar{n})) = \prod_{1\leq j_1<j_2\leq d} (n_{j_2} - n_{j_1}) = \prod_{1\leq j_1<j_2\leq d} (2j_2 - 2j_1)$$
$$= 2^{\binom{d}{2}} \prod_{1\leq j_1<j_2\leq d} (j_2 - j_1) = 2^{\binom{d}{2}} \prod_{i=1}^{d-1} i!.$$

This shows that $\det(B(\bar{n})) = 2^{\binom{d}{2}}$. $\square$

Lemma 14. *Let p be an odd prime. Given $0 \le k \le t$, there exists a multilinear polynomial $f \in \mathbb{F}_p[x_1, \ldots, x_t]$ of degree at most $d = \lfloor k/2 \rfloor$ such that for every $a \in \{0,1\}^t$*

$$f(a) = \begin{cases} 1\,, & \text{if } |a| = k \\ 0\,, & \text{if } |a| < k \text{ and } (|a| = k \bmod 2)\,. \end{cases}$$

Proof. We look for f of the form $f = \sum_{j=0}^{d} c_j S_{j,t}$ where $S_{j,t}$ is the elementary symmetric polynomial $S_{j,t} = \sum_{|A|=j} \prod_{i \in A} x_i$. Given $a \in \{0,1\}^t$,

$$f(a) = \sum_{j=0}^{d} c_j \binom{|a|}{j} \bmod p\,.$$

We are therefore looking for a solution of the linear system

$$B(\bar{n})\,(c_0 \ldots, c_d)^t = (0, \ldots, 0, 1)^t\ ,$$

where $\bar{n} = (0, 2, \ldots, 2d)$, if k is even, and $\bar{n} = (1, 3, \ldots, 2d+1)$, if k is odd. By the previous lemma, $B(\bar{n})$ is invertible over $\mathbb{F}_p$ and such a solution exists. $\square$

Proof of Lemma 9. Let $\mathbb{F}$ be a field of characteristic $p \neq 2$ containing a subfield $\mathbb{F}'$ such that every element of $\mathbb{F}'$ is a sum of r squares in $\mathbb{F}$. If $p = 0$, $\mathbb{F}'$ contains $\mathbb{Q}$ and if $p > 2$, $\mathbb{F}'$ contains $\mathbb{F}_p$. Let f be the polynomial given by Claim 10 or Lemma 14 with coefficients from $\mathbb{F}'$. Since every element of $\mathbb{F}'$ is a sum of r squares in $\mathbb{F}$, we can write

$$f(x_1, \ldots, x_t) = \sum_{C \in \mathcal{C}} a_C \prod_{i \in C} x_i\,,$$

where $\mathcal{C}$ is a multiset of $s \le r\binom{t}{\le d}$ subsets of $[t]$, and $a_C \in \mathbb{F}'$ has a square root $a_C^{\frac{1}{2}}$ in $\mathbb{F}$. For $A \in \binom{[t]}{k}$, let $\xi(A) \in \mathbb{F}^s$ be a vector whose coordinates are indexed by elements C of $\mathcal{C}$ so that

$$\xi(A)_C = \begin{cases} a_C^{\frac{1}{2}}\,, & \text{if } C \subseteq A \\ 0\,, & \text{if } C \not\subseteq A\,. \end{cases}$$

This gives a (k,t)-parity representation over $\mathbb{F}$. $\square$

4.2 Comments

An improvement on the dimension of parity representation in Lemma 9, if possible, will lead to an improvement in Theorem 1. However, this dimension cannot be too small:

Remark 15. *If k is even, every (k,t)-parity representation must have dimension at least $s = \binom{\lfloor t/2 \rfloor}{k/2}$ over any field. This is because there exists a family $\mathcal{A}$ of k-element subsets of $[t]$ whose pairwise intersection is even, and $|\mathcal{A}| = s$. The map ξ must assign linearly independent vectors to elements of $\mathcal{A}$. Similarly for k odd.*

On the other hand, Lemma 9 can sometimes be improved. $\binom{t}{\leq \lfloor k/2 \rfloor}$ can be replaced with $\binom{t}{\leq \lfloor t-k/2 \rfloor}$ which gives a smaller bound if if $k > t/2$. This is because we can work with complements of $A \in \binom{[t]}{k}$ instead. Another improvement is possible in odd characteristic for specific choices of k:

Remark 16. *If p is odd and $k = 2p^\ell - 1$, there is a (k,t)-parity representation of dimension $\binom{t}{\lfloor k/2 \rfloor}$ over $\mathbb{F}_p$. It follows from Lucas' theorem that in this case, f in Lemma 14 can be taken simply as the elementary symmetric polynomial of degree $\lfloor k/2 \rfloor$. This polynomial has only $\binom{t}{\lfloor k/2 \rfloor}$ monomials.*

The notion of (k,t)-parity representation can be restated in the language of *orthonormal representations* of graphs of Lovász [16]. Given a graph G with vertex set V, its orthonormal representation is a map $\xi(V) :\to \mathbb{F}^s$ such that for every $u, v \in V$

$$\begin{aligned} \langle \xi(u), \xi(u) \rangle &= 1\,, \\ \langle \xi(u), \xi(v) \rangle &= 0\,, \text{ if } u \neq v \text{ are not adjacent in } G. \end{aligned}$$

In this language, (k,t)-parity representation is an orthonormal representation of the following combinatorial Knesser-type graph $G_{k,t}$: vertices of $G_{k,t}$ are k-element subsets of $[t]$. There is an edge between u and v iff $|u \cap v| \neq k \bmod 2$. Orthogonal representations of related graphs have been studied by Haviv in [8, 7].

5 Modifications and extensions

5.1 A sum of bilinear products

Theorem 1 implies:

Theorem 17. *Over any field, there exists $s \leq O(n^{1.62})$ and bilinear $f_1, \dots, f_{2s}$ such that*

$$(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2) = f_1^2 + \dots + f_s^2 - (f_{s+1}^2 + \dots + f_{2s}^2)\,. \tag{8}$$

Proof. If $\mathbb{F}$ contains a square root of -1, Theorem 1 applies. Otherwise consider the field extension $\mathbb{F}^* = \mathbb{F}[\sqrt{-1}]$. Then we can express $(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2)$ as $f_1^2 + \dots + f_s^2$ over $\mathbb{F}^*$. Writing $f_k = g_k + \sqrt{-1}h_k$ where g_k and h_k have coefficients in $\mathbb{F}$ gives $(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2) = \sum_{k=1}^s (g_k^2 - h_k^2)$. □

From the point of view of arithmetic complexity, it is more natural to consider identities of the form

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = f_1 f_1' + \cdots + f_s f_s', \tag{9}$$

where $f_1, \ldots, f_s$ and $f_1', \ldots, f_s'$ are bilinear forms. This is because a non-commutative circuit computing ID_n leads to an identity of this form. This quantity is referred to as *bilinear complexity* in [11]. An upper bound on s in (9) can be inferred from Theorem 17. A direct proof was presented in [10].

Remark 18. *In characteristic different from two, we have* $ff' = \left(\frac{f+f'}{2}\right)^2 - \left(\frac{f-f'}{2}\right)^2$*, which allows to rewrite (9) as (8). In turn, we can express (8) as a sum of squares provided* -1 *is a sum of squares in* $\mathbb{F}$*. We conclude that, first, Theorem 17 implies Theorem 1 and, second, it is sufficient to consider the more general bilinear identities (9).*

5.2 A tensor product construction

We now outline an alternative construction of non-trivial sum-of-squares identities. While it gives different types of identities, it does not seem to give better bounds asymptotically.

Instead of the products of anticommuting matrices e_A, one can take the *tensor* product of matrices satisfying Hurwitz-Radon conditions (4). Namely, given such matrices $H_1, \ldots, H_m \in \mathbb{F}^{n \times s}$, and $a \in [m]^\ell$, let

$$H_a := H_{a_1} \otimes H_{a_2} \cdots \otimes H_{a_\ell}.$$

Observe that every H_a satisfies $H_a H_a^t = I_{n^\ell}$ and that

$$H_a H_b^t + H_b H_a^t = 0,$$

whenever a and b have *odd* Hamming distance (i.e., they differ in an odd number of coordinates). As in Lemma 9, we can find a map $\xi : [m]^\ell \to \mathbb{C}^s$ with $s \leq (4m)^{\ell/2}$ such that

$$\begin{aligned} \langle \xi(a), \xi(a) \rangle &= 1, \\ \langle \xi(a), \xi(b) \rangle &= 0, \text{ if } a \neq b \text{ have even Hamming distance.} \end{aligned}$$

This gives for every ℓ

$$\sigma_{\mathbb{C}}(n^\ell, m^\ell) \leq \sigma_{\mathbb{C}}(n, m)^\ell (4m)^{\ell/2}$$

For example, starting with $\sigma_{\mathbb{C}}(8, 8) = 8$, we have

$$\sigma_{\mathbb{C}}(8^\ell, 8^\ell) \leq 8^{11\ell/6}.$$

6 Non-commutative complexity of related polynomials

In this section, we prove Theorem 3. The main component is a construction of a subquadratic circuit for ID_n (Theorem 25). The upper bound for $S_{4,n}$ and $\mathsf{perm}_{4,n}$ follows by reduction to ID_n (Corollary 27).

Commutative and non-commutative arithmetic circuits In Section 2, we introduced non-commutative arithmetic circuits. Given non-commutative polynomials $f_1, \dots, f_m$ over a field $\mathbb{F}$, we will denote $\mathsf{size}_{\mathbb{F}}^{(nc)}(f_1, \dots, f_m)$ the size of a smallest non-commutative arithmetic circuit over $\mathbb{F}$ simultaneously computing $f_1, \dots, f_m$, namely, such that every f_i is computed by *some* gate in the circuit. A *commutative* arithmetic circuit is the more common model for computing polynomials in the commutative ring $\mathbb{F}[x_1, \dots, x_n]$. It is defined similarly as non-commutative arithmetic circuit, except that the order of multiplication is irrelevant. The commutative complexity will be denoted $\mathsf{size}_{\mathbb{F}}^{(c)}$. Given a non-commutative polynomial f, let $f^{(c)}$ be the same polynomial f in which the variables are viewed as commutative. This means

$$\mathsf{size}_{\mathbb{F}}^{(c)}(f^{(c)}) \leq \mathsf{size}_{\mathbb{F}}^{(nc)}(f)\,.$$

We will drop the subscript $\mathbb{F}$ if the field is arbitrary or clear from the context.

Proof outline of Theorem 3 for ID_n We first show that in order to bound the *non-commutative* complexity of ID_n, it is sufficient to construct a *commutative* sum-of-squares identity (1) with few squares such that the bilinear forms $f_1, \dots, f_s$ can be simultaneously computed by a small arithmetic circuit. This is the content of Lemma 22. The proof is a more elaborate version of a similar argument in [11].

In the ideal world, we would proceed to show that the bilinear forms constructed in Theorem 1 are indeed computable by a circuit of subquadratic size. A related question is to estimate the tensor rank of an associated tensor (which amounts to counting the number of non-scalar multiplications in a circuit). The tensor obtained in Theorem 1 is simple enough to describe but we do not know how to bound its rank. The construction from Section 5.2 is easier to analyze. A conditional upper bound on tensor rank can be obtained assuming Strassen's asymptotic rank conjecture [21], but it is unclear how to obtain it unconditionally.

Fortunately, this issue can be avoided completely by using Theorem 1 in a black-box fashion. Suppose that we can write $(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2)$ as $\sum_{j=1}^s f_j(\bar{x}, \bar{y})^2$ where $f_j(\bar{x}, \bar{y})$ have some unknown complexity. Introducing m copies of the y variables we obtain a new sum-of-squares identity

$$(\sum_{i=1}^n x_i^2)(\sum_{i\in[n],t\in[m]} y_{i,t}^2) = \sum_{j\in[s],t\in[m]} f_j(\bar{x}, \bar{y}_t)^2\,.$$

This is wasteful in terms of the number of squares but less so in terms of their complexity. Computing m copies of $f_1(\bar{x},\bar{y}),\dots,f_s(\bar{x},\bar{y})$ can be done efficiently using fast matrix multiplication. If m is large enough, the complexity of the initial polynomials is irrelevant and the resulting complexity is determined by matrix multiplication only. This argument gives a worse upper bound for ID_n than the previous bound on $\sigma(n)$, but still a subquadratic one. The connection with matrix multiplication is further discussed in Section 6.3

6.1 Some facts about bilinear forms

We now overview some basic facts about bilinear forms. The one non-trivial ingredient is a result of Baur and Strassen [2] on computing partial derivatives of a polynomial. We will need the following simple version of their result:

Lemma 19. *[Baur-Strassen] Let $f_1,\dots,f_r$ be (commutative) polynomials not depending on variables $z_1,\dots,z_r$. Then $\mathsf{size}^{(c)}(f_1,\dots,f_r)\leq O(\mathsf{size}^{(c)}(\sum_{i=1}^r f_i z_i))$.*

In the non-commutative setting, a *bilinear form* in variables $\bar{x}=(x_1,\dots,x_n)$ and $\bar{y}=(y_1,\dots,y_m)$ will be taken as a polynomial of the form $\sum_{i,j} a_{i,j}x_iy_j$.

Lemma 20. *Let $f_1,\dots,f_r$ be non-commutative bilinear forms and $f:=\sum_{k=1}^r f_kz_k$. Then*

$$\begin{aligned}\mathsf{size}^{(nc)}(f_1,\dots,f_r) &\leq O(\mathsf{size}^{(c)}(f_1^{(c)},\dots,f_r^{(c)}))\,,\\ \mathsf{size}^{(nc)}(f) &\leq O(\mathsf{size}^{(c)}(f^{(c)})\,.\end{aligned}$$

Proof. Given a commutative circuit Ψ computing $f_1^{(c)},\dots,f_r^{(c)}$, we can, by increasing its size by a constant factor, assume that it is homogeneous. That is, every gate computes a homogeneous polynomial of degree at most two (this is a standard construction, see, e.g. [3, 15]). Given a linear function h in variables $\bar{x}$, $\bar{y}$, we can write $h=h_X+h_Y$ where h_X and h_Y depend on variables $\bar{x}$ only or $\bar{y}$ only, respectively. In the circuit Ψ, we can first split every gate v computing a linear function h into two gates v_X, v_Y computing h_X and h_Y. Second, a product gate $v\cdot v'$ computing a product of linear functions can be replaced by the non-commutative product $v_X\cdot v'_Y+v'_X\cdot v_Y$.

If f has a commutative arithmetic circuit of size s then $f_1,\dots,f_r$ can be simultaneously computed by a commutative circuit of size $O(s)$ by Lemma 19 and hence by a non-commutative circuit of linear size as well. This gives $\mathsf{size}^{(nc)}(f)\leq O(r+s)$. Without loss of generality, we can assume that all f_k's are non-zero so that $r\leq s$ which gives the required bound. □

Remark 21. *The lemma implies that the non-commutative complexities of*

$$\sum_{i,j,k} a_{i,j,k}x_iy_jz_k\,,\quad\text{and}\quad \sum_{i,j,k} a_{i,j,k}x_iz_ky_j$$

differ by a constant factor only.

6.2 From sum-of-squares to a circuit for ID_n

Let $\gamma_{\mathbb{F}}(n,m)$ denote the smallest k such that there exist bilinear $f_1,\dots,f_s$ which satisfy the *commutative* identity (1) and can be simultaneously computed by a *commutative* arithmetic circuit of size k.

Lemma 22. *Let $\mathbb{F}$ be a field of characteristic different from. Let $\mathbb{F}^*$ be the smallest field extension of $\mathbb{F}$ containing a square root of -1. Then $\mathsf{size}^{(nc)}_{\mathbb{F}}(\mathsf{ID}_{n,m}) = O(\gamma_{\mathbb{F}^*}(n,m))$.*

Proof. We will assume that $\mathbb{F}$ contains a square root of -1 so that $\mathbb{F}^* = \mathbb{F}$. If this is not the case, we can view an element of $\mathbb{F}^\star = \mathbb{F}[\sqrt{-1}]$ as a pair of elements of $\mathbb{F}$ and simulate a computation over $\mathbb{F}^\star$ in $\mathbb{F}$ (cf. [13]). This gives $\gamma_{\mathbb{F}}(n,m) \leq O(\gamma_{\mathbb{F}^*}(n,m))$.

Let $f = \sum_{i,j} a_{i,j}x_iy_j$ be a commutative bilinear form and z a new variable. Define the following non-commutative polynomials

$$f^{xy} := \sum_{i,j} a_{i,j}x_iy_j\,,\ f^{yx} := \sum_{i,j} a_{i,j}y_jx_i\,,$$
$$f\star z := \sum_{i,j} a_{i,j}x_izy_j\,,\quad f^{[2]} := \frac{1}{2}(f^{xy}f^{xy} + f\star f^{yx})\,.$$

$f^{[2]}$ mimics the commutative polynomial f^2 in the following sense:

Claim. *Given $i,i' \in [n]$ and $j,j' \in [m]$, let $c(i,j,i',j')$ and $\bar{c}(i,j,i',j')$ denote the coefficient of $x_iy_jx_{i'}y_{j'}$ in f^2 and $f^{[2]}$, respectively. Then $\bar{c}(i,j,i',j') = \lambda(i,j,i',j')c(i,j,i',j')$, where*

$$\lambda(i,j,i',j') = \begin{cases} 1\,, & \text{if } i=i', j=j'\,,\\ \frac{1}{2}\,, & \text{if } i=i',\ j\neq j'\,,\ \text{or vice versa,}\\ \frac{1}{4}\,, & \text{if } i\neq i'\,, j\neq j'\,. \end{cases}$$

Proof of the claim. By definition of $f^{[2]}$, the coefficient of $x_iy_jx_{i'}y_{j'}$ in $f^{[2]}$ is

$$\bar{c}(i,j,i',j') = \frac{1}{2}(a_{i,j}a_{i',j'} + a_{i,j'}a_{i',j})\,. \tag{10}$$

On the other hand, considering possible ways of factoring $x_iy_jx_{i'}y_{j'}$ into bilinear monomials, its coefficient in f^2 equals

$$c(i,j,i',j') = \begin{cases} a_{i,j}^2\,, & \text{if } i=i', j=j'\\ 2a_{i,j}a_{i,j'}\,, & \text{if } i=i',\ j\neq j'\\ 2a_{i',j}a_{i',j}\,, & \text{if } i\neq i',\ j=j'\\ 2(a_{i,j}a_{i',j'} + a_{i,j'}a_{i',j})\,, & \text{if } i\neq i'\,, j\neq j' \end{cases}\,.$$

Comparing this with (10) gives the required statement. □

Suppose that $\gamma_{\mathbb{F}}(n,m) = r$. We can then write

$$(\sum_{i\in[n]} x_i^2)(\sum_{j\in[m]} y_j^2) = \sum_{k\in[s]} a_k f_k^2\,,$$

where $f_1,\dots,f_s$ are distinct commutative bilinear forms with $\mathsf{size}^{(c)}(f_1,\dots,f_s) = r$ and $a_1,\dots,a_s \in \mathbb{F}$. Since $\mathsf{ID}_{n,m}^{(c)}$, when viewed as a commutative polynomial, equals $(\sum_i x_i^2)(\sum y_j^2)$, the above Claim shows that

$$\mathsf{ID}_{n,m} = \sum_{k\in[s]} a_k f_k^{[2]}\,.$$

We now estimate the complexity of $\sum_{k=1}^{s} a_k f_k^{[2]}$. Introducing new variables $z_1,\dots,z_s$, let G be the polynomial

$$G(z_1,\dots,z_s) := \sum_{k\in[s]} a_k f_k \star z_k\,.$$

Viewed as a commutative polynomial, $G^{(c)}$ equals $\sum_{k\in[s]} a_k f_k z_k$. Since $f_1,\dots,f_s$ can be simultaneously computed by a circuit of size r, $G^{(c)}$ has a commutative circuit of size linear in r. By Lemma 20, the same holds for the non-commutative polynomial G. Writing

$$\sum_{k\in[s]} a_k f_k^{[2]} = \sum_{k\in[s]} \frac{1}{2}(a_k f_k^{xy} f_k^{xy} + G(f_1^{yx},\dots,f_s^{yx}))\,.$$

gives a circuit of size $O(r)$. $\square$

Remark 23. *The opposite inequality $\gamma_{\mathbb{F}^*}(n,m) \le O(\mathsf{size}_{\mathbb{F}}^{(nc)}(\mathsf{ID}_{n,m}))$ also holds.*

Proof sketch. Let ψ be a non-commutative circuit computing ID_n. As shown in [11], we can assume it has the following additional structure: it is homogeneous and every gate computing a degree-two polynomial computes either a non-commutative bilinear form in $\bar{x}$ and $\bar{y}$, or a bilinear form in $\bar{y}$ and $\bar{x}$. We now view ψ as a *commutative* circuit computing $(\sum_i x_i^2)(\sum_j y_j^2)$ with the additional property that every degree-two gate computes a bilinear form. For every degree-two gate v computing f_v, introduce a new variable z_v. For every product gate $w = u \cdot v$ with v computing a polynomial of degree 2 and u of degree ≥ 1, replace w with $u \cdot z_v$. Let F be the polynomial computed by this new circuit. F is multilinear in the variables z_v and

$$(\sum_i x_i^2)(\sum_j y_j^2) = \sum_v f_v \partial_{z_v} F\,.$$

The bilinear forms f_v are simultaneously computed by the circuit ψ itself. $\partial_{z_v} F$ have a small circuit using Lemma 20. The polynomials $\partial_{z_v} F$ are not necessarily bilinear but their "bilinear parts" can be efficiently computed. This gives

$(\sum_i x_i^2)(\sum_j y_j^2) = \sum_v f_v f_v'$ where f_v, f_v' are bilinear and can be simultaneously computed by a commutative circuit of size $O(\mathsf{size}_{\mathbb{F}}^{(nc)}(\mathsf{ID}_n))$. Finally, $\sum_v f_v f_v'$ can be converted to a sum-of-squares identity over $\mathbb{F}^*$ as in Remark 18. $\square$

Let $\omega(r)$ be the exponent of rectangular matrix multiplication capturing the complexity of multiplying $n \times n^r$ matrix by an $n^r \times n$ matrix. It is the least (infimum) value such that the matrix product can be computed by a (commutative) arithmetic circuit of size $O(n^{\omega(r)+\epsilon})$ for every $\epsilon > 0$. We will use the estimates on $\omega(r)$ as given by le Gall and Urrutia [5].

Lemma 24. *Let $r \geq 2$ be an integer and $\delta \geq 0$. Let $Q(\bar{x}, \bar{y})$ be a set of $O(n^{1+\delta})$ bilinear forms in (commuting) variables $\bar{x} = (x_1, \dots, x_n)$, $\bar{y} = (y_1, \dots, y_n)$. Let $\bar{y}_1, \dots, \bar{y}_m$ be distinct copies of $\bar{y}$ with $m := n^r$. Then $Q(\bar{x}, \bar{y}_1), \dots, Q(\bar{x}, \bar{y}_m)$ can be simultaneously computed by an arithmetic circuit of size $n^{\omega(r)+\delta+o(1)}$.*

Proof. Splitting $Q(\bar{x}, \bar{y})$ into $O(n^\delta)$ sets of size n, it is sufficient to prove the statement for $Q(\bar{x}, \bar{y})$ consisting of n bilinear forms $f_1(\bar{x}, \bar{y}), \dots, f_n(\bar{x}, \bar{y})$. Let f be the trilinear polynomial $\sum_{k=1}^n f_k z_k$ in variables $\bar{x}, \bar{y}$ and $\bar{z}$. Introduce new variables $y_{i,t}$, $z_{t,i}$, $t \in [m]$, $i \in [n]$. If $f = \sum_{i,j,k \in [n]} a_{i,j,k} x_i y_j z_k$, let

$$f^\star := \sum_{i,j,k \in [n]} a_{i,j,k} x_i \sum_{t \in [m]} y_{j,t} z_{t,k} \,.$$

This guarantees that

$$f^\star = \sum_{k \in [n], t \in [m]} f_k(\bar{x}, \bar{y}_t) z_{t,k} \,.$$

By Lemma 20, it is sufficient to estimate the complexity of $f^\star$. The polynomials $\sum_{t \in [m]} y_{j,t} z_{t,k}$, $i, k \in [n]$, can be simultaneously computed in size $O(n^{\omega(r)+\epsilon})$. Each of the n^2 linear functions $\sum_{k \in [n]} a_{i,j,k} x_k$, $i, j \in [n]$, can be computed by a circuit of size $O(n)$. Hence the complexity of $f^\star$ is $O(n^{\omega(r)+\epsilon} + n^3)$. If $r \geq 2$ then $\omega(r) \geq 3$ and the cubic term can be omitted. $\square$

Theorem 25. *Over a field of characteristic different from two, $\mathsf{size}^{(nc)}(\mathsf{ID}_n) \leq O(n^c)$ with $c < 1.96$.*

Proof. Using Lemma 22, it is enough to estimate $\gamma_{\mathbb{F}}(n, n)$ under the assumption that $\mathbb{F}$ contains a square root of -1. By Theorem 1, we can write

$$(\sum_{i=1}^n x_i^2)(\sum_{i=1}^n y_i^2) = \sum_{j=1}^s f_j(\bar{x}, \bar{y})^2 \,,$$

with $s = O(n^{1+\delta})$ and $\delta < 0.616$. Introducing $m = n^3$ copies of the $\bar{y}$ variables we obtain a new sum-of-squares identity

$$(\sum_{i=1}^n x_i^2)(\sum_{i \in [n], t \in [m]} y_{i,t}^2) = \sum_{j \in [s], t \in [m]} f_j(\bar{x}, \bar{y}_t)^2 \,.$$

From the previous lemma, we obtain, for every $\epsilon > 0$,

$$\gamma_{\mathbb{F}}(n, n^4) = O(n^{\omega(3)+\delta+\epsilon})\,.$$

Duplicating the $\bar{x}$ variables n^3 times gives $\gamma_{\mathbb{F}}(n^4, n^4) \leq n^3 \gamma_{\mathbb{F}}(n, n^4)$. Hence, $\gamma_{\mathbb{F}}(n^4, n^4) = O(n^{3+\omega(3)+\delta+\epsilon})$ and

$$\gamma_{\mathbb{F}}(n, n) \leq n^{\frac{3+\omega(3)+\delta}{4}+o(1)}\,.$$

In [5], it is shown that $\omega(3) < 4.1997$ which gives $\gamma_{\mathbb{F}}(n, n) = O(n^{1.954})$. $\square$

6.3 Comments

The numerical value of the exponent in Theorem 25 can be slightly improved. First, we can analyze the complexity of the bilinear forms constructed in Theorem 1 and, second, use asymmetric bounds on $\sigma(n, n^k)$ for a suitable k. However, these improvements are too minuscule to justify the more complicated proof.

The complexity of matrix multiplication enters the picture quite naturally. Consider Euler's four-square identity

$$(x_1^2 + \cdots + x_4^2)(y_1^2 + \cdots + y_4^2) = f_1^2 + \cdots + f_4^2\,.$$

Here, the bilinear map $f = (f_1, \ldots, f_4)$ can be interpreted as computing the product of two quaternions so that

$$(x_1 + x_2 i + x_3 j + x_4 k)(y_1 + y_2 i + y_3 j + y_4 k) = f_1 + f_2 i + f_3 j + f_4 k\,,$$

where i, j, k satisfy the familiar properties $i^2, j^2, k^2 = -1$, $k = ij = -ji$. The basis elements $1, i, j, k$ can be represented in terms of 2×2 complex matrices $1_{\mathbb{C}}, i_{\mathbb{C}}, j_{\mathbb{C}}, k_{\mathbb{C}}$. These are linearly independent and form a basis of the space of 2×2 complex matrices. This means that over $\mathbb{C}$, the number of non-scalar multiplications required to compute the map f is *exactly the same* as the number of non-scalar multiplications needed to multiply two 2×2 matrices.

A similar connection holds between the complexity of multiplying two $2^n \times 2^n$ matrices and the complexity of multiplication in the second Clifford algebra $\mathrm{C}L_{2n+1}$. An element of $\mathrm{C}L_m$ is of the form $\sum_A x_A e_A$ where i) A ranges over even subsets of $[m]$, and ii) if $i_1 < \cdots < i_k$, $e_{\{i_1,\ldots,i_k\}} = e_{i_1} e_{i_2} \cdots e_{i_k}$ where $e_1, \ldots, e_m$ satisfy $e_i^2 = 1$ and $e_i e_j = -e_j e_i$ whenever $i \neq j$. Hence, $\mathrm{C}L_2$ corresponds to $\mathbb{C}$ and $\mathrm{C}L_3$ to quaternions. An alternative way of obtaining a subquadratic sum-of-squares identity is as follows: in the first step, compute the product of two elements of $\mathrm{C}L_m$ by means of a bilinear map f. This gives a sum-of-squares identity for $m \leq 3$ but no longer works for a larger m. In the second step, tweak the map f by using the parity representation as in Theorem 11. In terms of the arithmetic complexity of the resulting map, already the first step is equivalent to matrix multiplication.

6.4 An application to elementary symmetric polynomials

Recall the non-commutative polynomials $S_{k,n}$ and $\mathsf{perm}_{k,n}$ from Section 2. As follows from Theorem 7.1 in [11], they have almost the same complexity:

$$\mathsf{size}^{(nc)}(S_{k,n}) \le \mathsf{size}^{(nc)}(\mathsf{perm}_{k,n}) \le O(k^3 \mathsf{size}^{(nc)}(S_{k,n}))\,. \tag{11}$$

This means that we can focus just on the polynomial $S_{k,n}$.

Proposition 26. *Over any field,* $\mathsf{size}^{(nc)}(S_{2,n}, S_{3,n}) \le O(n)$ *and* $\mathsf{size}^{(nc)}(S_{4,n}) \le O(\mathsf{size}^{(nc)}(\mathsf{ID}_n))$.

Proof. Let $p_k := \sum_{i=1}^n x_i^k$. Omitting the subscript n in $S_{k,n}$,

$$S_2 = p_1^2 - p_2\,,$$

giving a circuit of a linear size for S_2. We can write

$$S_3 = p_1 S_2 - p_2 p_1 - \sum_i x_i p_1 x_i + 2p_3\,.$$

Note that $\sum x_i p_1 x_i$ has a linear-sized circuit: we can first compute $\sum x_i z x_i$ and then substitute p_1 for z. This gives a linear circuit for S_3.

Let $\mathsf{ID}^* := \sum_{i,j\in[n]} x_i x_j x_i x_j$. Hence, ID^* is obtained by identifying y_i with x_i, $i \in [n]$, in ID_n. We can write

$$S_4 = p_1 S_3 - \sum_{i,j,k} x_i^2 x_j x_k - \sum_{i,j,k} x_i x_j x_i x_k - \sum_{i,j,k} x_i x_j x_k x_i\,,$$

where i, j, k range ever *distinct* elements of $[n]$. The complexity of $p_1 S_3$ is linear. We claim that the other summands have either a linear circuit size, or are easily computable from ID^*. We can write

$$\sum_{i,j,k} x_i^2 x_j x_k = p_2 S_2 - p_3 p_1 - \sum_i x_i^2 p_1 x_i + 2p_4\,,$$
$$\sum_{i,j,k} x_i x_j x_k x_i = \sum_i x_i S_2 x_i - \sum_i x_i^2 p_1 x_i - \sum_i x_i p_1 x_i^2 + 2p_4\,.$$

giving a circuit of size $O(n)$. Similarly,

$$\sum_{i,j,k} x_i x_j x_i x_k = \sum_i x_i p_1 x_i p_1 - \mathsf{ID}^* - \sum_i x_i p_1 x_i^2 - p_3 p_1 + 2p_4\,.$$

and the complexity is bounded by $\mathsf{size}^{(nc)}(\mathsf{ID}^*) + O(n)$. $\square$

Corollary 27. *Assume that the underlying field has characteristic different from two. There exists a constant* $c < 1.96$ *such that* $\mathsf{size}^{(nc)}(S_{4,n}) = O(n^c)$ *and* $\mathsf{size}^{(nc)}(S_{k,n}) = O(n^{k-4+c})$ *for every fixed* $k \ge 4$. *Similarly for* $\mathsf{perm}_{k,n}$.

Proof. If $k = 4$, the bound on $S_{k,n}$ follows from Proposition 26 and Theorem 25. For $k > 4$, the identity

$$S_{k,n}(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i S_{k-1,n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

gives $\mathsf{size}^{(nc)}(S_{k,n}) \leq O(\mathsf{size}^{(nc)}(n^{k-4} S_{4,n}))$. The part for $\mathsf{perm}_{4,n}$ follows from (11). $\square$

Remark 28. *A non-commutative polynomial $f(x_1, \ldots, x_n)$ is* symmetric *if $f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n)$ holds for every permutation σ of $[n]$. As in Proposition 26, it can be show that $\mathsf{size}^{(nc)}(f) \leq O(\mathsf{size}^{(nc)}(\mathsf{ID}^*))$ holds for any non-commutative symmetric n-variate polynomial of degree four. In other words, $\sum_{i,j \in [n]} x_i x_j x_i x_j$ is a symmetric polynomial of degree four with the largest non-commutative complexity.*

7 Open problems

Let Even_t denote the set of even-sized subsets of $[t]$. A map $\xi : \text{Even}_t \to \mathbb{F}^s$ will be called a *t-parity representation* of dimension s if for every $A, B \in \text{Even}_t$

$$\begin{aligned} \langle \xi(A), \xi(A) \rangle &= 1\,, \\ \langle \xi(A), \xi(B) \rangle &= 0\,, \text{ if } A \neq B \text{ and } |A \cap B| \text{ is even.} \end{aligned}$$

Problem 1. *Over $\mathbb{C}$, does there exist a t-parity representation of dimension $2^{(0.5+o(1))t}$?*

If this were the case, we could improve the bound of Theorem 1 to $\sigma_{\mathbb{C}}(n, n) \leq n^{1.5+o(1)}$. A more surprising consequence would be that

$$\sigma_{\mathbb{C}}(n, n^2) \leq n^{2+o(1)}\,.$$

The constant 0.5 in Problem 1 cannot be improved: since there exists a family of $2^{\lfloor t/2 \rfloor}$ subsets of $[t]$ with pairwise even intersection, every t-parity representation must have dimension at least $2^{\lfloor t/2 \rfloor}$ (cf. Remark 15). On the other hand, Lemma 9 implies that there exists a t-parity representation of dimension at most $2^{(H(0.25)+o(1))t} < 2^{0.82t}$.

Our results do not apply to sum-of-squares composition formulas over the real numbers. Since $\mathbb{R}$ is one of the most natural choices of the underlying field, it is desirable to extend the construction in this direction. This motivates the following:

Problem 2. *Over $\mathbb{R}$, does there exist a t-parity representation of dimension $O(2^{t(1-\epsilon)})$ with $\epsilon > 0$?*

While the sum-of-squares problem trivializes in a field of characteristic two, the construction of a subquadratic circuit for ID_n does not work in this case.

Problem 3. *Over a field of characteristic two, can ID_n be computed by a non-commutative circuit of size $O(n^{2-\epsilon})$ with $\epsilon > 0$?*

References

[1] V. Arvind, A. Chatterjee, R. Datta, and P. Mukhopadhyay. On explicit branching programs for the rectangular determinant and permanent polynomials. *Chic. J. Theor. Comput. Sci.*, 2020, 2019.

[2] W. Baur and V. Strassen. The complexity of partial derivatives. *Theoretical Computer Science*, 22:317–330, 1983.

[3] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic complexity theory*, volume 315 of *A series of comprehensive studies in mathematics*. Springer, 1997.

[4] M. L. Carmosino, R. Impagliazzo, S. Lovett, and I. Mihajlin. Hardness amplification for non-commutative arithmetic circuits. In *Proceedings of the 33rd Computational Complexity Conference*, CCC '18, 2018.

[5] F. Le Gall and F. Urrutia. Improved rectangular matrix multiplication using powers of the Coppersmith-Winograd tensor. In *ACM-SIAM Symposium on Discrete Algorithms*, 2017.

[6] A. Geramita and N. Pullman. A theorem of Hurwitz and Radon and orthogonal projective modules. *Proceedings of The American Mathematical Society*, 42:51–51, 01 1974.

[7] I. Haviv. On minrank and the Lovász Theta function. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2018.

[8] I. Haviv. Topological bounds on the dimension of orthogonal representations of graphs. *European Journal of Combinatorics*, 81:84–97, 2019.

[9] P. Hrubeš. On families of anticommuting matrices. *Linear Algebra and Applications*, 493:494–507, 2016.

[10] P. Hrubeš. A subquadratic upper bound on sum-of-squares composition formulas. *ECCC*, 2024.

[11] P. Hrubeš, A. Wigderson, and A. Yehudayoff. Non-commutative circuits and the sum of squares problem. *J. Amer. Math. Soc.*, 24:871–898, 2011.

[12] P. Hrubeš, A. Wigderson, and A. Yehudayoff. An asymptotic bound on the composition number of integer sums of squares formulas. *Canadian Mathematical Bulletin*, 56:70–79, 2013.

[13] P. Hrubeš and A. Yehudayoff. Arithmetic complexity in ring extensions. *Theory of Computing*, 7:119–129, 2011.

[14] A. Hurwitz. Über die Komposition der quadratischen Formen von beliebigvielen Variabeln. *Nach. Ges. der Wiss. Göttingen*, pages 309–316, 1898.

[15] S. Berkowitz L. Valiant, S. Skyum and C. Rackoff. Fast parallel computation of polynomials using few processors. *Siam J. Comp.*, 12:641–644, 1983.

[16] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.

[17] N. Nisan. Lower bounds for non-commutative computation. In *Proceeding of the 23th STOC*, pages 410–418, 1991.

[18] J. Radon. Lineare scharen orthogonalen Matrizen. *Abh. Math. Sem. Univ. Hamburg*, 1(2-14), 1922.

[19] D. B. Shapiro. Quadratic forms and similarities. *Bull. Amer. Math. Soc.*, 81(6), 1975.

[20] D. B. Shapiro. *Compositions of quadratic forms.* De Gruyter expositions in mathematics 33, 2000.

[21] V. Strassen. Algebra and complexity. In *Progr. Math.*, volume 120, pages 429–446. 1994.

[22] V. Vassilevska and R.Williams. Finding, minimizing, and counting weighted subgraphs. *SIAM Journal on Computing*, 42:455–464, 05 2009.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il