


Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $\text{VNC}^1 \neq \text{VNP}$

Benjamin Rossman  

Duke University, Durham, NC, USA

Davidson Zhu  

Duke University, Durham, NC, USA

Abstract

The *sum-of-squares (SoS) complexity* of a d -multiquadratic polynomial f (quadratic in each of d blocks of n variables) is the minimum s such that $f = \sum_{i=1}^s g_i^2$ with each g_i d -multilinear. In the case $d = 2$, Hrubeš, Wigderson and Yehudayoff [13] showed that an $n^{1+\Omega(1)}$ lower bound on the SoS complexity of explicit biquadratic polynomials implies an exponential lower bound for non-commutative arithmetic circuits. In this paper, we establish an analogous connection between general *multiquadratic sum-of-squares* and *commutative arithmetic formulas*. Specifically, we show that an $n^{d-o(\log d)}$ lower bound on the SoS complexity of explicit d -multiquadratic polynomials, for any $d = d(n)$ with $\omega(1) \leq d(n) \leq O(\frac{\log n}{\log \log n})$, would separate the algebraic complexity classes VNC^1 and VNP .

2012 ACM Subject Classification Theory of computation \rightarrow Algebraic complexity theory

Keywords and phrases sum-of-squares, arithmetic formulas

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.113

Related Version *Full Version*: <https://arxiv.org/abs/2512.01227>

1 Introduction

Computational complexity theory aims to categorize computational problems by the amount of time and resources they need. The algebraic complexity classes proposed by Leslie Valiant [31], on the other hand, were created for understanding the tractability of polynomials, as opposed to Boolean functions. Arithmetic circuits, branching programs and formulas are defined as analogs of their Boolean counterparts.

Algebraic complexity classes $\text{VF} (= \text{VNC}^1) \subseteq \text{VBP} \subseteq \text{VP} (= \text{VNC}^2) \subseteq \text{VNP}$ consist of sequences of $\text{poly}(n)$ -degree n -variate polynomials that are computable by $\text{poly}(n)$ -size arithmetic formulas, branching programs, circuits and projections of circuits, respectively. As with the analogous Boolean classes ($\text{NC}^1 \subseteq \text{NL} \subseteq \text{P} \subseteq \text{NP}$), these four algebraic classes are believed to be distinct. However, it remains a major open problem even to separate VNC^1 from VNP , despite significant progress on lower bounds for formulas, branching programs and circuits in various restricted settings: low-depth [2, 17, 18], multilinear [9, 20, 23], set-multilinear [3, 16, 25], and non-commutative [6, 19, 30] (to mention just a few representative works).

1.1 Biquadratic sum-of-squares and non-commutative circuits

Let $f \in \mathbb{F}[X_1, \dots, X_n, Y_1, \dots, Y_n]$ be a biquadratic polynomial over a field \mathbb{F} with monomials of the form $X_{i_1} X_{i_2} Y_{j_1} Y_{j_2}$. The (*biquadratic*) *sum-of-squares complexity* of f , denoted $\text{SoS}(f)$, is the minimum number s of bilinear polynomials g_1, \dots, g_s such that $f = g_1^2 + \dots + g_s^2$. When \mathbb{F} is algebraically closed, we have $\text{SoS}(f) \leq n^2$ for all f , while a dimension-counting argument shows that $\text{SoS}(f) = \Omega(n^2)$ for almost all f .



© Benjamin Rossman and Davidson Zhu;

licensed under Creative Commons License CC-BY 4.0

17th Innovations in Theoretical Computer Science Conference (ITCS 2026).

Editor: Shubhangi Saraf; Article No. 113; pp. 113:1–113:22

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Sum-of-squares complexity has been intensively studied for the separable biquadratic polynomial

$$q_n := (X_1^2 + \dots + X_n^2) \cdot (Y_1^2 + \dots + Y_n^2).$$

Over \mathbb{R} and \mathbb{C} , trivial bound are given by

$$n \leq \text{SoS}_{\mathbb{C}}(q_n) \leq \text{SoS}_{\mathbb{R}}(q_n) \leq n^2.$$

A classical theorem of Hurwitz from 1898 shows that $\text{SoS}_{\mathbb{R}}(q_n) = n$ iff $n \in \{1, 2, 4, 8\}$ [15]. An upper bound $\text{SoS}_{\mathbb{R}}(q_n) = O(\frac{n^2}{\log n})$ is given by the classical Radon-Hurwitz construction (see [22, 28]). Asymptotically, this remains the strongest known upper bound over \mathbb{R} . It was also the strongest known bound over \mathbb{C} , until a recent breakthrough of Hrubeš [14]:

► **Theorem 1.1** ([14]). *There exists a sum-of-squares decomposition of q_n over \mathbb{C} with just $O(n^{1.62})$ terms, that is, $\text{SoS}_{\mathbb{C}}(q_n) = O(n^{1.62})$.*

As for lower bounds, Adams and Atiyah [1] showed $\text{SoS}_{\mathbb{C}}(q_n) \geq 2^{\lceil \log_2 n \rceil}$; in particular, $\text{SoS}_{\mathbb{C}}(q_n) \geq 2n - 2$ infinitely often. Slightly better bounds have been obtained for specific values of n (and for general fields of characteristic $\neq 2$) using techniques from algebraic K-theory [7]. However, no lower bound better than $\text{SoS}_{\mathbb{C}}(q_n) = \Omega(n)$ is currently known.

A striking result of Hrubeš, Wigderson and Yehudayoff [13] explains the difficulty of improving this lower bound. They show that proving a sufficiently super-linear lower bound on $\text{SoS}_{\mathbb{C}}(q_n)$ would have a dramatic consequence in non-commutative algebraic circuit complexity:

► **Theorem 1.2** ([13]). *Let $f = (f_n)$ be any explicit sequence¹ of biquadratic polynomials over \mathbb{C} . If $\text{SoS}_{\mathbb{C}}(f) = \Omega(n^{1+\varepsilon})$ for any $\varepsilon > 0$, then the permanent PERM_n requires exponential size non-commutative circuits; in particular, this implies a separation of complexity classes VP_{nc} and VNP_{nc} (the non-commutative algebraic analogues of P and NP). Moreover, a lower bound $\text{SoS}_{\mathbb{C}}(f) = \Omega(n \log^5 n)$ suffices for the separation $\text{VP}_{\text{nc}} \neq \text{VNP}_{\text{nc}}$.*

Notice that Hrubeš's upper bound $\text{SoS}_{\mathbb{C}}(q_n) = O(n^{1.62})$ does not rule out the possibility of separating VP_{nc} and VNP_{nc} via an $\Omega(n^{1+\varepsilon})$ lower bound. (We remark that Theorems 1.1 and 1.2 hold more generally over any algebraically closed field, but are not known to extend to \mathbb{R} ; see the discussion in §1.4.)

1.2 Multiquadratic sum-of-squares and commutative set-multilinear formulas

The main theorem of this paper shows that strong enough lower bounds for *multiquadratic sum-of-squares* imply super-polynomial lower bounds for *commutative arithmetic formulas*.

Let f be a d -multiquadratic polynomial over an algebraically closed field, that is, a linear combination of monomials $X_{i_1}^{(1)} X_{j_1}^{(1)} \dots X_{i_d}^{(d)} X_{j_d}^{(d)}$ where $i_1, j_1, \dots, i_d, j_d \in [n]$. As in the biquadratic setting ($d = 2$), the (*multiquadratic*) *sum-of-squares complexity* of f , denoted $\text{SoS}(f)$, is defined as the minimum number s of d -multilinear polynomials g_1, \dots, g_s such that $f = g_1^2 + \dots + g_s^2$. Similar to the biquadratic setting, we have $\text{SoS}(f) \leq O(n^d)$ for all f , and $\text{SoS}(f) = \Omega((n/2)^d)$ for almost all f (Prop. 3.17).

¹ The sequence $f = (f_n)$ is *explicit* if the coefficient in f_n of each monomial $X_{i_1} X_{i_2} Y_{j_1} Y_{j_2}$ is computable in $\text{polylog}(n)$ time given n and indices $i_1, i_2, j_1, j_2 \in [n]$.

Our main result – stated informally below and formally in the next subsection – shows that a nearly optimal lower bound on the SoS complexity of explicit d -multiquadratic polynomials would separate the algebraic complexity classes VNC^1 and VNP .

► **Theorem 1.3 (informal).** *Let $f = (f_n)$ be any explicit sequence of d -multiquadratic polynomials over an algebraically closed or finite field of characteristic $\neq 2$. If $\text{SoS}(f) = n^{d-o(\log d)}$, then this yields an $n^{\Omega(\log d)}$ lower bound on the set-multilinear formula size of associated $d + 1$ -multilinear polynomials $\tilde{f} = (\tilde{f}_n)$. If in addition $d = d(n)$ satisfies $\omega(1) \leq d(n) \leq O(\frac{\log n}{\log \log n})$, we get a separation $\text{VNC}^1 \neq \text{VNP}$ (via a lemma of Raz [24]).*

While Theorem 1.3 is analogous to the result of Hrubeš, Wigderson and Yehudayoff [13], it is closer in nature to a well-known theorem of Raz [24] linking *tensor rank* to set-multilinear formula lower bounds (see the discussion in §1.4).

1.3 Partial transpose rank of $n^d \times n^d$ matrices

► **Definition 1.4.** *Let $\mathcal{X} = \{X_{i_1}^{(1)}\}_{i_1=1}^n \cup \{X_{i_2}^{(2)}\}_{i_2=1}^n \cup \dots \cup \{X_{i_d}^{(d)}\}_{i_d=1}^n$ be d disjoint blocks of n variables. A polynomial $P(\mathcal{X})$ is d -multiquadratic if each monomial of $P(\mathcal{X})$ has total degree exactly 2 in each block.*

Every $n^d \times n^d$ matrix M gives rise to a d -multiquadratic polynomial defined by

$$Q_M := \sum_{\vec{i}, \vec{j} \in [n]^d} M_{\vec{i}, \vec{j}} \cdot X_{i_1}^{(1)} X_{j_1}^{(1)} \dots X_{i_d}^{(d)} X_{j_d}^{(d)}.$$

Conversely, for fields of characteristic $\neq 2$, we get a bijective correspondence between d -multiquadratic polynomials (with d blocks of n variables) and $n^d \times n^d$ matrices satisfying $M = M^{\top_1} = \dots = M^{\top_d}$, where \top_k is the k -th partial transpose operation which swaps row-index i_k and column-index j_k . We refer to such matrices as *fully symmetric*.

The actual main theorem of this paper concerns a generalization of SoS complexity called *partial transpose rank* – or *PT-rank* for short – which is a complexity measure on general $n^d \times n^d$ matrices (not just fully symmetric ones). We say that an $n^d \times n^d$ matrix P is *PT-basic* if P^{\top_κ} has rank 1 for any $\kappa \subseteq [d]$, where \top_κ is the composition of (commuting) operations \top_k for $k \in \kappa$. We then define the *PT-rank* of any $n^d \times n^d$ matrix M as the minimum number r of PT-basic matrices P_1, \dots, P_r such that $M = P_1 + \dots + P_r$. This complexity measure is closely related to SoS complexity: for fully symmetric matrices M (satisfying $M = M^{\top_k}$ for all $k \in [d]$), we show that $\text{PT-rank}(M)$ and $\text{SoS}(Q_M)$ are equivalent to an $O(2^d)$ factor over any algebraically closed or finite field of characteristic $\neq 2$ (Prop. 3.15).

Our main theorem relates the PT-rank of an arbitrary $n^d \times n^d$ matrix M with the set-multilinear formula size of an associated multilinear polynomial of degree $d + 1$ in $dn + (d - 1)n^2 + dn$ variables:

$$\tilde{Q}_M := \sum_{\vec{i}, \vec{j} \in [n]^d} M_{\vec{i}, \vec{j}} \cdot \tilde{X}_{i_1}^{(0)} \tilde{X}_{j_1, i_2}^{(1)} \dots \tilde{X}_{j_{d-1}, i_d}^{(d-1)} \tilde{X}_{j_d}^{(d)}.$$

► **Theorem 1.5 (Main theorem).** *For every $n^d \times n^d$ matrix M (over any field), the set-multilinear formula size of \tilde{Q}_M is at least $\text{PT-rank}(M)/n^{d-\log d+1}$.*

Via a standard set-multilinearization lemma of Raz [24], we get the following corollary:

► **Corollary 1.6.** *Suppose that $\text{PT-rank}(M) = n^{d-o(\log d)}$ where M is any explicit sequence of $n^d \times n^d$ matrices with $\omega(1) \leq d(n) \leq O(\frac{\log n}{\log \log n})$. Then \tilde{Q}_M has arithmetic formula size at least $n^{(1-o(1)) \log d}$ and, hence, $\text{VNC}^1 \neq \text{VNP}$.*

(Theorem 1.3 on SoS complexity follows directly from Theorem 1.5 and Corollary 1.6. For a d -multiquadratic polynomial $f = Q_M$, the multilinear polynomial \tilde{f} of Theorem 1.3 is given by \tilde{Q}_M .)

Since almost all M have $\text{PT-rank} \geq (\frac{1}{2} - o(1))n^d$ (Prop. 3.16), finding an explicit M with $\text{PT-rank}(M) = n^{d-o(\log d)}$ is a classic problem of “finding hay in a haystack”. By finding a very explicit M (with \tilde{Q}_M in VP rather than VNP), Corollary 1.6 could conceivably be used to show $\text{VNC}^1 \neq \text{VP}$. On the flipside, an additional result of this paper (Theorem 6.6) rules out lower bounds, over fields of characteristic $\neq 2$, whenever \tilde{Q}_M belongs to the class VBP_{ord} of set-multilinear polynomials computable by a polynomial-size ordered branching program. This negative result leverages the recent upper bound $\text{SoS}_{\mathbb{C}}(q_n) = O(n^{1.62})$ of Hrubeš [14] (Theorem 1.1). We also show that $\text{PT-rank}(M)$ is at most $n^{O(d^{0.7})}$ when M is a Kronecker product of d $n \times n$ matrices over any fields of characteristic $\neq 2$ (Theorem 6.3); this observation rules out certain explicit $n^d \times n^d$ matrices as candidates for having $\text{PT-rank} n^{d-o(\log d)}$.

So far as we know, over $\text{GF}(2)$, it is possible that the $n^d \times n^d$ identity matrix I_{n^d} has $\text{PT-rank} n^{d-o(\log d)}$ or even exactly n^d . If so, by Theorem 1.5 this implies an $n^{(1-o(1)) \log d}$ lower bound on the formula size of iterated matrix multiplication $\text{IMM}_{n,d+1}$ ($= \tilde{Q}_{I_{n^d}}$), which would separate classes VNC^1 and VBP over $\text{GF}(2)$.

1.4 Related work

Following the influential paper of Hrubeš, Wigderson and Yehudayoff [13], subsequent work by various authors [3, 5, 8] has found additional reductions between various sum-of-squares complexity measures and algebraic circuit lower bounds. Most recently, Dutta, Saxena and Thierauf [8] showed that an $d^{0.5+\Omega(1)}$ lower bound on the support-sum of any explicit degree- d univariate polynomial would separate VP and VNP. However, none of the prior work has explored the d -multiquadratic setting, nor implications specific to arithmetic formulas.

More closely related to our main theorem is the well-known theorem of Raz [24] that an $n^{d-o(d)}$ lower bound on the tensor rank of any explicit d -multilinear polynomial g (i.e. order- d tensor $[n]^d \rightarrow \mathbb{F}$) implies an $n^{\Omega(\log d)}$ lower bound on the set-multilinear formula size of the same polynomial g . The further implication $\text{VNC}^1 \neq \text{VNP}$ when $\omega(1) \leq d(n) \leq O(\frac{\log n}{\log \log n})$ follows from an additional lemma in [24], which we also invoke above.

We remark that a lower bound $\text{SoS}(f) = n^{d-o(\log n)}$ for an explicit d -multiquadratic polynomial f (i.e. fully symmetric $n^d \times n^d$ matrix) is conceivable without having to break the longstanding $O(n^{\lfloor d/2 \rfloor})$ tensor rank barrier. Indeed, Theorem 1.3 appears to evade the known barriers for rank-based lower bound methods identified by Efremenko *et al* [10] and Garg *et al* [11].

1.5 Outline of the paper

The rest of the paper is organized as follows: §2 presents key definitions pertaining to tensors, matrices, and arithmetic formulas. §3 formally introduces *partial transpose rank* and explores its properties. §4 proves a weaker (but easier and illustrative) version of our main theorem – with respect to *non-commutative* set-multilinear formulas. §5 proves the main theorem – with respect to *commutative* set-multilinear formulas. §6 discusses upper bounds for PT-rank and candidate explicit hard matrices.

2 Preliminaries

Throughout this paper, n, d are arbitrary positive integers, where we regard n as a growing parameter and $d = d(n)$ as a function of n .

We write $[n]$ for the set $\{1, \dots, n\}$ and $\log n$ for the base-2 logarithm of n . We write $\langle \cdot, \cdot \rangle : [n]^2 \rightarrow [n^2]$ for the pairing bijection defined by $\langle i, j \rangle := (i-1)n + j$.

2.1 The underlying field

Throughout this paper, \mathbb{F} is an arbitrary field. We will explicitly state whenever an assumption on \mathbb{F} is required. In particular, our main theorem (Theorem 1.5) and other results concerning PT-rank work for completely general fields (including characteristic 2).

The only results that depend on \mathbb{F} concern the relationship between PT-rank and SoS complexity (Lemma 3.15) and the consequence of our main theorem stated in the terms of SoS complexity (Theorem 1.3). Even here, our results apply to a large class of fields of characteristic $\neq 2$, including all algebraically closed fields and finite fields.

2.2 Matrices and tensors

In the Introduction, our main theorem was stated first in terms of SoS complexity of d -multiquadratic polynomials, then more generally in terms of PT-rank of $n^d \times n^d$ matrices. These rank measures were linked to the algebraic complexity of a third kind of object: $d+1$ -multilinear polynomials. It is convenient to regard these objects as different flattenings of a tensor $[n]^{2d} \rightarrow \mathbb{F}$ to tensors of format $[n^2]^d$ or $[n^d]^2$ or $[n] \times [n^2]^{d-1} \times [n]$, respectively.

► **Definition 2.1 (Tensors).** An order d tensor is a function $A : [n_1] \times \dots \times [n_d] \rightarrow \mathbb{F}$ where n_1, \dots, n_d are nonnegative integers.² We shall mainly (but not exclusively) consider tensors $[n]^d \rightarrow \mathbb{F}$ where $n_1 = \dots = n_d = n$. More generally, we consider tensors $A : [n]^D \rightarrow \mathbb{F}$ with coordinates indexed by a nonempty finite set D .³

For tensors $A : [n]^d \rightarrow \mathbb{F}$ and $B : [n]^e \rightarrow \mathbb{F}$, their tensor product is denoted by $A \otimes B : [n]^{d+e} \rightarrow \mathbb{F}$. For tensors $A : [n]^D \rightarrow \mathbb{F}$ and $B : [n]^E \rightarrow \mathbb{F}$ with D and E disjoint, their tensor product is denoted by $A \otimes B : [n]^{D \cup E} \rightarrow \mathbb{F}$.

The tensor rank of a tensor $A : [n_1] \times \dots \times [n_d] \rightarrow \mathbb{F}$, denoted $\text{tensor-rank}(A)$, is the minimum number r such that A admits a decomposition $A = \sum_{j=1}^r (A_{1,j} \otimes \dots \otimes A_{d,j})$ for some collection of order 1 tensors $A_{i,j} : [n_i] \rightarrow \mathbb{F}$.⁴

► **Definition 2.2 (Matrices).** A matrix is an order 2 tensor $M : [m_1] \times [m_2] \rightarrow \mathbb{F}$, also written as $M \in \mathbb{F}^{m_1 \times m_2}$. The rank of M is denoted by $\text{rank}(M)$; note that this is a special case tensor rank. The transpose of M is denoted by $M^\top \in \mathbb{F}^{m_2 \times m_1}$. (We use a bolder font \top to distinguish the full transpose from partial transpose operations \top_κ on $n^d \times n^d$ matrices introduced in Def. 3.1.)

² All definitions and claims in this paper can be rephrased in a basis-free manner, that is, regarding A as a multilinear function $\mathbb{F}^{n_1} \times \dots \times \mathbb{F}^{n_d} \rightarrow \mathbb{F}$ without reference to any specific bases for vector spaces \mathbb{F}^{n_i} .

³ Specifically, in §5 we consider tensors where D is the directed edge set of an arbitrary subgraph of Path_d .

⁴ We define *tensor rank* solely for the sake of comparing our main result (Theorem 1.5) with Raz's theorem shows that an $n^{d-o(d)}$ lower bound on tensor rank implies $\text{VNC}^1 \neq \text{VP}$ [24], as discussed in §1.4. Elsewhere in this paper, we only every consider the matrix rank of flattened tensors.

For matrices $M : [m_1] \times [m_2] \rightarrow \mathbb{F}$ and $N : [n_1] \times [n_2] \rightarrow \mathbb{F}$, their Kronecker product is denoted by $M \boxtimes N : [m_1 n_1] \times [m_2 n_2] \rightarrow \mathbb{F}$ (not to be confused with the tensor product $M \otimes N : [m_1] \times [m_2] \times [n_1] \times [n_2] \rightarrow \mathbb{F}$.)

The $n \times n$ identity matrix is denoted by I_n . We write $I_n^{\otimes d}$ for the order $2d$ tensor $I_n \otimes \cdots \otimes I_n$ (d times) (not to be confused with the $n^d \times n^d$ matrix $I_{n^d} = I_n \boxtimes \cdots \boxtimes I_n$, which is a tensor of order 2).

When M is an $n^d \times n^d$ matrix, we index its rows and columns by d -tuples in $[n]^d$. That is, we express the entries of M as $M_{(i_1, \dots, i_d), (j_1, \dots, j_d)}$.

► **Definition 2.3** (Matrix flattening of a tensor). For a tensor $A : [n]^D \rightarrow \mathbb{F}$ and partition $I \sqcup J = D$, we denote $\text{Mat}_{I,J}(A)$ for the flattened matrix

$$\text{Mat}_{I,J}(A) : [n^{|I|}] \times [n^{|J|}] \rightarrow \mathbb{F}.$$

2.3 Set-multilinear formula size

There is a one-to-one correspondence between order d tensors $A : [n_1] \times \cdots \times [n_d] \rightarrow \mathbb{F}$ and set-multilinear polynomials $P_A \in \mathbb{F}_{\text{sm}}[\mathcal{X}]$ in variables $\mathcal{X} = \{X_a^{(i)} : i \in [d], a \in [n_i]\}$. Namely, A describes the coefficients of P_A :

$$P_A(\mathcal{X}) = \sum_{(a_1, \dots, a_d) \in [n_1] \times \cdots \times [n_d]} A_{a_1, \dots, a_d} X_{a_1}^{(1)} \cdots X_{a_d}^{(d)}.$$

For any nonempty finite set D , we have a similar correspondence between tensors $A : [n]^D \rightarrow \mathbb{F}$ and set-multilinear polynomial $P_A \in \mathbb{F}_{\text{sm}}[\mathcal{X}]$ in variables $\mathcal{X} = \{X_a^{(i)} : i \in D, a \in [n]\}$.

► **Definition 2.4** (Set-multilinear formula size). The set-multilinear formula size of a tensor $A : [n]^D \rightarrow \mathbb{F}$, denoted $\mathcal{L}_{\text{sm}}(A)$, is the minimum number of leaves in a syntactically set-multilinear arithmetic formula that computes the associated polynomial P_A . This may also be defined directly by induction:

- If $|D| = 1$, then $\mathcal{L}_{\text{sm}}(A)$ is 1 if A is non-zero anywhere and 0 otherwise.
- If $|D| \geq 2$, then $\mathcal{L}_{\text{sm}}(A)$ is the minimum value of $\sum_i (\mathcal{L}_{\text{sm}}(B_i) + \mathcal{L}_{\text{sm}}(C_i))$ over indexed families $\{(E_i, F_i, B_i, C_i)\}_i$ (where i runs over an arbitrary, unnamed index set) such that E_i, F_i are disjoint nonempty sets satisfying $E_i \cup F_i = D$ and $B_i : [n]^{E_i} \rightarrow \mathbb{F}$ and $C_i : [n]^{F_i} \rightarrow \mathbb{F}$ are tensors satisfying $A = \sum_i B_i \otimes C_i$.

For further background on algebraic complexity measures, see [4, 26, 29].

2.4 $n^d \times n^d$ matrices and their shifted tensors

Recall that our main theorem, Theorem 1.5, relates $\text{PT-rank}(M)$ to the set-multilinear formula size of a related polynomial. This polynomial is denoted \tilde{Q}_M in the statement of Theorem 1.5. However, it is convenient to speak of the associated tensor, which we shall refer to as the “shifted tensor” of M .

► **Definition 2.5** (The shifted tensor of an $n^d \times n^d$ matrix M). For any $n^d \times n^d$ matrix M , we defined the reshaped tensor $\text{ShiftedTensor}(M) : [n^2]^{d+1} \rightarrow \mathbb{F}$ as follows. Recall that $\langle \cdot, \cdot \rangle$ is a bijection $[n]^2 \rightarrow [n^2]$. For all $p, i_1, j_1, \dots, i_d, j_d, q \in [n]$, we have

$$\text{ShiftedTensor}(M)(\langle p, i_1 \rangle, \langle j_1, i_2 \rangle, \dots, \langle j_{d-1}, i_d \rangle, \langle j_d, q \rangle) := \mathbb{1}\{p = q = 1\} \cdot M_{(i_1, \dots, i_d), (j_1, \dots, j_d)}.$$

Note that the set-multilinear polynomial $P_{\text{ShiftedTensor}(M)}$ is precisely \tilde{Q}_M of Theorem 1.5.

► **Definition 2.6** (Iterated Matrix Multiplication). *An important example of a “shifted tensor” is the Iterated Matrix Multiplication tensor $\text{IMM}_{n,d} : [n^2]^d \rightarrow \mathbb{F}$ defined by*

$$\text{IMM}_{n,d}(\langle a_1, b_1 \rangle, \dots, \langle a_d, b_d \rangle) := \mathbb{1}[(b_1, \dots, b_{d-1}) = (a_2, \dots, a_d)].$$

The familiar corresponding set-multilinear polynomial describes the sum of entries in the product of d symbolic $n \times n$ matrices:

$$P_{\text{IMM}_{n,d}}(\mathcal{X}) = \sum_{a_0, \dots, a_d \in [n]} X_{\langle a_0, a_1 \rangle}^{(1)} X_{\langle a_1, a_2 \rangle}^{(2)} \cdots X_{\langle a_{d-1}, a_d \rangle}^{(d)}.$$

Note that $\text{IMM}_{n,d} = \text{ShiftedTensor}(I_{n^{d-1}})$, that is, $\text{IMM}_{n,d}$ is the shifted tensor of the $n^{d-1} \times n^{d-1}$ identity matrix.

3 Partial transpose rank

The partial transpose is a key concept in quantum information theory, particularly in studying entanglement and separability of quantum states [12, 21]. In this section, we introduce the *partial transpose rank* of an $n^d \times n^d$ matrix. To the best of our knowledge, this notion has not been explicitly studied before. As we show, PT rank generalizes SoS complexity (for fully symmetric matrices over algebraically closed or finite fields of characteristic $\neq 2$).

We begin by defining the *partial transpose* operations on $n^d \times n^d$ matrices.

► **Definition 3.1** (Partial transpose). *Let M be an $n^d \times n^d$ matrix.*

■ *For $k \in [d]$, the k -th partial transpose of M , denoted $M^{\top k}$, is the $n^d \times n^d$ matrix defined by*

$$M_{(i_1, \dots, i_d), (j_1, \dots, j_d)}^{\top k} := M_{(i_1, \dots, i_{k-1}, j_k, i_{k+1}, \dots, i_d), (j_1, \dots, j_{k-1}, i_k, j_{k+1}, \dots, j_d)}.$$

■ *For $\kappa \subseteq [d]$, the κ -partial transpose of M is the matrix $M^{\top \kappa} := M^{\top k_1 \cdots \top k_c}$ where k_1, \dots, k_c are the distinct elements of κ (in any order).*

Note that $M^{\top \emptyset} = M$ and $M^{\top [d]} = M^{\top}$ (the usual transpose of M). Also note that

$$M^{\top \kappa_1 \top \kappa_2} = M^{\top \kappa_2 \top \kappa_1} = M^{\top \kappa_1 \Delta \kappa_2}$$

where $\kappa_1 \Delta \kappa_2 = (\kappa_1 \setminus \kappa_2) \cup (\kappa_2 \setminus \kappa_1)$ is the symmetric difference of κ_1 and κ_2 . In particular, the transpose of $M^{\top \kappa}$ is $M^{\top \bar{\kappa}}$ where $\bar{\kappa} := [d] \setminus \kappa$ is the complement of κ . If M is symmetric (i.e. $M^{\top} = M$), it follows that $M^{\top \kappa} = M^{\top \bar{\kappa}}$.

► **Definition 3.2** (Fully symmetric matrices). *We say that an $n^d \times n^d$ matrix M is fully symmetric if $M^{\top k} = M$ for all $k \in [d]$ (equivalently: if $M^{\top \kappa} = M$ for all $\kappa \subseteq [d]$).*

► **Example 3.3.** Let $d = 2$ and $n = 3$. Suppose M is a partitioned $3^2 \times 3^2$ matrix with 3×3 blocks M_{11}, \dots, M_{33} . Then

$$M^{\top \emptyset} = M = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, \quad M^{\top \{1\}} = M^{\top 1} = \begin{bmatrix} M_{11}^{\top} & M_{12}^{\top} & M_{13}^{\top} \\ M_{21}^{\top} & M_{22}^{\top} & M_{23}^{\top} \\ M_{31}^{\top} & M_{32}^{\top} & M_{33}^{\top} \end{bmatrix},$$

$$M^{\top \{2\}} = M^{\top 2} = \begin{bmatrix} M_{11} & M_{21} & M_{31} \\ M_{12} & M_{22} & M_{32} \\ M_{13} & M_{23} & M_{33} \end{bmatrix}, \quad M^{\top \{1,2\}} = M^{\top} = \begin{bmatrix} M_{11}^{\top} & M_{21}^{\top} & M_{31}^{\top} \\ M_{12}^{\top} & M_{22}^{\top} & M_{32}^{\top} \\ M_{13}^{\top} & M_{23}^{\top} & M_{33}^{\top} \end{bmatrix}.$$

When $d = 2$ (as in this example), the operation \top_1 is also known as the *block transpose*, since it transposes the individual blocks within a partitioned $n^2 \times n^2$ matrix.

- **Definition 3.4** (Partial transpose rank). *Let M be an $n^d \times n^d$ matrix.*
- M is PT-basic if there exists $\kappa \subseteq [d]$ such that $\text{rank}(M^{\top \kappa}) = 1$.
 - The PT-rank of M is the minimum number of PT-basic matrices that sum to M .

The next lemma gives an alternative characterization of PT-rank.

- **Lemma 3.5.** *The PT-rank of an $n^d \times n^d$ matrix M equals the minimum value of $\sum_{\kappa \subseteq [d-1]} \text{rank}(N_{\kappa}^{\top \kappa})$ over all decompositions $M = \sum_{\kappa \subseteq [d-1]} N_{\kappa}$.*

Proof. Let r be the PT-rank of M and let B_1, \dots, B_r be PT-basic matrices that sum to M . For each B_i , there exists $\kappa_i \subseteq [d]$ with $\text{rank}(B_i^{\top \kappa_i}) = 1$. Without loss of generality, we may choose $\kappa_i \subseteq [d-1]$, since $\text{rank}(B_i^{\top \kappa_i}) = \text{rank}((B_i^{\top \kappa_i})^{\top}) = \text{rank}(B_i^{\top \overline{\kappa_i}})$. Now define $N_{\kappa} := \sum_{i \in [r]: \kappa_i = \kappa} B_i$. ◀

The next two examples illustrate PT-rank in the case $d = 2$.

- **Example 3.6.** Let $d = 2$ and $n = 2$. The $2^2 \times 2^2$ identity matrix has PT-rank 2, since it is not PT-basic, but is a sum of PT-basic matrices:

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] = \left[\begin{array}{cc|cc} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{array} \right] + \left[\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]^{\top_1}$$

3.1 Properties of PT-rank

Unlike the full transpose ($\kappa = [d]$), the partial transpose operation can increase matrix rank when $\emptyset \subsetneq \kappa \subsetneq [d]$. (See full version of the paper for proofs of lemmas in this section.)

- **Lemma 3.7.** *For every $n^d \times n^d$ matrix M and $\kappa \subseteq [d]$, we have*

$$\text{rank}(M^{\top \kappa}) \leq n^{2 \min\{|\kappa|, d-|\kappa|\}} \text{rank}(M).$$

Moreover, this inequality is best possible in general.

The partial transpose also behaves differently than the full transpose with respect to products of matrices. For two $n^d \times n^d$ matrices P and M , we have $(PM)^{\top} = M^{\top} P^{\top}$; additionally, $\text{rank}((PM)^{\top}) = \text{rank}(M^{\top})$ whenever P is invertible. With respect to the partial transpose, we get analogous equations only in the case that P (or M) is a Kronecker product of d $n \times n$ matrices.

- **Lemma 3.8.** *Let P and M be $n^d \times n^d$ matrices such that $P = B_1 \boxtimes \dots \boxtimes B_d$ for $n \times n$ matrices B_1, \dots, B_d . Then for all $\kappa \subseteq [d]$,*

$$(PM)^{\top \kappa} = LM^{\top \kappa}R \quad \text{where } L = \bigotimes_{k=1}^d \begin{cases} I_n & \text{if } k \in \kappa \\ B_k & \text{otherwise} \end{cases} \quad \text{and } R = \bigotimes_{k=1}^d \begin{cases} B_k^{\top} & \text{if } k \in \kappa \\ I_n & \text{otherwise.} \end{cases}$$

Moreover, if B_1, \dots, B_d are nonsingular then $\text{rank}((PM)^{\top \kappa}) = \text{rank}(M^{\top \kappa})$.

- **Corollary 3.9.** *PT-rank is invariant under the action of $\text{GL}(n)^d$ on $n^d \times n^d$ matrices.*

Corollary 3.9 shows that PT-rank generalizes to a complexity measure on multilinear functions $V_1 \times \dots \times V_d \rightarrow V_1 \times \dots \times V_d$ (equivalently: on tensors in $\bigotimes_{c=1}^d (V_c \otimes V_c^*)$) for any n -dimensional \mathbb{F} -vector spaces V_1, \dots, V_d , independent of a choice of bases.

Note that the PT-rank of a matrix depends on its assumed dimension. For example, the PT-rank of an $n^4 \times n^4$ matrix is different from its PT-rank treated as a $\widehat{n}^2 \times \widehat{n}^2$ matrix where $\widehat{n} = n^2$, as a different set of partial transposes are allowed. When there is ambiguity, we write $\text{PT-rank}_{[n]^4}(M)$ or $\text{PT-rank}_{[n^2]^2}(M)$ to distinguish these two cases.

► **Lemma 3.10.** $\text{PT-rank}_{[n]^{pq}}(M) \leq \text{PT-rank}_{[n^p]^q}(M)$ whenever $d = pq$ with $p, q \geq 1$.

Proof. Write M for the $n^{pq} \times n^{pq}$ matrix and N for the equivalent $\hat{n}^q \times \hat{n}^q$ matrix with $\hat{n} = n^p$. For $\lambda \subseteq [q]$, note that

$$\text{rank}(N^{\top \lambda}) = \text{rank}(M^{\top \kappa}) \quad \text{where } \kappa = \{k + (l-1)p : k \in [p], l \in [\lambda]\}.$$

The claimed inequality follows from this fact and the characterization of PT-rank given by Lemma 3.5. ◀

3.2 PT-rank and SoS complexity

We now explain the close relationship between PT-rank of $n^d \times n^d$ matrices and SoS complexity of d -multiquadratic forms.

► **Definition 3.11** (Multiquadratic polynomial Q_M). For an $n_1 \cdots n_d \times n_1 \cdots n_d$ matrix M , let Q_M denote the d -multiquadratic polynomial

$$Q_M := \sum_{\vec{i}, \vec{j} \in [n_1] \times \cdots \times [n_d]} M_{\vec{i}, \vec{j}} \cdot X_{i_1}^{(1)} X_{j_1}^{(1)} \cdots X_{i_d}^{(d)} X_{j_d}^{(d)}.$$

For example, the identity matrix I_{n^3} has corresponding 3-multiquadratic polynomial

$$q_{I_{n^3}} = \left(\sum_{i \in [n]} X_i^2 \right) \left(\sum_{i \in [n]} Y_i^2 \right) \left(\sum_{i \in [n]} Z_i^2 \right)$$

where we write X_i, Y_i, Z_i for $X_i^{(1)}, X_i^{(2)}, X_i^{(3)}$. (The biquadratic polynomial q_n in the Introduction is given by $q_{I_{n^2}}$ in this notation.)

► **Lemma 3.12.** Let \mathbb{F} be algebraically closed field of characteristic $\neq 2$. Then over \mathbb{F} , we have $\text{SoS}(Q_M) \leq \text{PT-rank}(M)$.

Proof. Suppose $\text{SoS}(Q_M) = m < \infty$. Then, there exist an $m \times n^d$ matrix L such that

$$Q_M = \sum_{\ell=1}^m \left(\sum_{\vec{i} \in [n]^d} L_{\ell, \vec{i}} \cdot X_{i_1}^{(1)} \cdots X_{i_d}^{(d)} \right)^2 = \sum_{\vec{i}, \vec{j} \in [n]^d} (L^{\top} L)_{\vec{i}, \vec{j}} \cdot X_{i_1}^{(1)} X_{j_1}^{(1)} \cdots X_{i_d}^{(d)} X_{j_d}^{(d)}.$$

Over an algebraically closed field, we have the familiar fact

$$\{\text{symmetric matrices of rank } \leq m\} = \{L^{\top} L : L \in \mathbb{F}^{m \times n^d}\}.$$

Therefore,

$$\text{SoS}(Q_M) = \min \left\{ \text{rank}(D) : \text{symmetric } D \text{ s.t. } Q_M = \sum_{\vec{i}, \vec{j} \in [n]^d} D_{\vec{i}, \vec{j}} \cdot X_{i_1}^{(1)} X_{j_1}^{(1)} \cdots X_{i_d}^{(d)} X_{j_d}^{(d)} \right\}.$$

Since variables commute, the righthand equation is equivalent to $M = \frac{1}{2^d} \sum_{\kappa \subseteq [d]} D^{\top \kappa}$. That is,

$$\text{SoS}(Q_M) \leq \min \left\{ \text{rank}(D) : M = \frac{1}{2^d} \sum_{\kappa \subseteq [d-1]} D^{\top \kappa} \right\}.$$

Given any decomposition $M = \sum_{\kappa \subseteq [d]} P_{\kappa}^{\top \kappa}$ with $\text{PT-rank}(M) = \sum_{\kappa} \text{rank}(P_{\kappa})$, let $D = \frac{1}{2^d} \sum_{\kappa} P_{\kappa}$. Then

$$\frac{1}{2^d} \sum_{\kappa} D^{\top \kappa} = \frac{1}{2^d} \sum_{\kappa} \sum_{\lambda} P_{\lambda}^{\top \kappa} = \frac{1}{2^d} \sum_{\kappa} \left(\sum_{\lambda} P_{\lambda}^{\top \lambda} \right)^{\top \kappa} = \frac{1}{2^d} \sum_{\kappa} M^{\top \kappa} = M.$$

Therefore, $\text{SoS}(Q_M) \leq \text{rank}(D) \leq \sum_{\kappa} \text{rank}(P_{\kappa}) = \text{PT-rank}(M)$. ◀

113:10 Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $\text{VNC}^1 \neq \text{VNP}$

► **Remark 3.13.** Over finite fields \mathbb{F} , the bound $\text{SoS}(Q_M) - 1 \leq \text{PT-rank}(M)$ can be proved using same ideas as above. The only difference is that following the result from [27], every symmetric matrix of rank m can be expressed as $L^\top L$ for $L \in \mathbb{F}^{(m+1) \times n^d}$, thus comes the additional 1.

As we show next, the bounds above are tight up to $O(2^d)$ factor for fully symmetric matrices.

► **Lemma 3.14.** *If M is a fully symmetric matrix over a field of characteristic $\neq 2$, then $\text{PT-rank}(M) \leq 2^{d-1} \text{SoS}(Q_M)$.*

Proof. Suppose $\text{SoS}(Q_M) = m < \infty$. Then there exists an $m \times n^d$ matrix L such that

$$Q_M = \sum_{\ell=1}^m \left(\sum_{\vec{i} \in [n]^d} L_{\ell, \vec{i}} \cdot X_{i_1}^{(1)} \cdots X_{i_d}^{(d)} \right)^2 = \sum_{\vec{i}, \vec{j} \in [n]^d} (L^\top L)_{\vec{i}, \vec{j}} \cdot X_{i_1}^{(1)} X_{j_1}^{(1)} \cdots X_{i_d}^{(d)} X_{j_d}^{(d)}.$$

Since these are commutative polynomials (in particular, $X_{i_k}^{(k)} X_{j_k}^{(k)} = X_{j_k}^{(k)} X_{i_k}^{(k)}$), it follows that

$$\sum_{\kappa \subseteq [d]} M^{\top \kappa} = \sum_{\kappa \subseteq [d]} (L^\top L)^{\top \kappa}.$$

Since M is fully symmetric and $L^\top L$ is symmetric, it follows that $2^{d-1} M = \sum_{\kappa \subseteq [d-1]} (L^\top L)^{\top \kappa}$. Therefore, we have $M = \sum_{\kappa \subseteq [d-1]} N_\kappa$ for matrices $N_\kappa := \frac{1}{2^{d-1}} (L^\top L)^{\top \kappa}$. Since $\text{rank}((N_\kappa)^{\top \kappa}) = \text{rank}(L^\top L) = s$ for each $\kappa \subseteq [d-1]$, we have $\text{PT-rank}(M) \leq 2^{d-1} s = 2^{d-1} \text{SoS}(Q_M)$. ◀

Lemmas 3.14 and 3.12 together imply:

► **Proposition 3.15.** *Let M be an $n^d \times n^d$ matrix over algebraically closed or finite field of characteristic $\neq 2$. Then $\text{SoS}(Q_M) - 1 \leq \text{PT-rank}(M) \leq 2^{d-1} \text{SoS}(Q_M)$.*

3.3 PT-rank of almost all matrices

Although not required for any other results in this paper, we conclude this section by observing that (asymptotically) almost all $n^d \times n^d$ matrices have PT-rank $\Omega(n^d)$, and (asymptotically) almost all fully symmetric $n^d \times n^d$ matrices have PT-rank $\Omega((n/2)^d)$. These results are established standard dimension-counting arguments. (See full paper for proofs.)

► **Proposition 3.16.** *Asymptotically almost all $n^d \times n^d$ matrices have PT-rank at least $(\frac{1}{2} - \varepsilon)n^d$ for every constant $\varepsilon > 0$.*

Since the set of fully symmetric $n^d \times n^d$ matrices has dimension $\geq n^{2d}/2^d$, a similar counting argument shows:

► **Proposition 3.17.** *Asymptotically almost all fully symmetric $n^d \times n^d$ matrices have PT-rank at least $(\frac{1}{2^{d+1}} - \varepsilon)n^d$ for any constant $\varepsilon > 0$.*

4 Non-commutative set-multilinear formula lower bounds from PT-rank

Recall the main theorem of this paper:

► **Theorem 1.5** (restated). *For every $n^d \times n^d$ matrix M , we have*

$$\mathcal{L}_{\text{sm}}(\text{ShiftedTensor}(M)) \geq \frac{\text{PT-rank}(M)}{n^{d-\log d+1}}.$$

As a helpful warm-up to the proof in §5, in this section we shall prove a strictly weaker – but slightly simpler – version of Theorem 1.5 that lower bounds the *non-commutative* set-multilinear formula size of $\text{ShiftedTensor}(M)$. We start with the definition:

► **Definition 4.1.** *The non-commutative set-multilinear formula size of a tensor $A : [n]^d \rightarrow \mathbb{F}$, denoted $\mathcal{L}_{\text{nc,sm}}(A)$, is defined inductively as follows:*

- *If $d = 1$, then $\mathcal{L}_{\text{nc,sm}}(A)$ equals 1 if A is non-zero at any point and 0 otherwise.*
- *If $d \geq 2$, then $\mathcal{L}_{\text{nc,sm}}(A)$ is the minimum value of $\sum_i (\mathcal{L}_{\text{nc,sm}}(B_i) + \mathcal{L}_{\text{nc,sm}}(C_i))$ over indexed families $\{(e_i, f_i, B_i, C_i)\}_i$ where e_i, f_i are positive integers satisfying $e_i + f_i = d$ and $B_i : [n]^{e_i} \rightarrow \mathbb{F}$ and $C_i : [n]^{f_i} \rightarrow \mathbb{F}$ are tensors such that $A = \sum_i B_i \otimes C_i$.*

In the remainder of this section, we prove the following:

► **Theorem 4.2** (Non-commutative version of Theorem 1.5). *For every $n^d \times n^d$ matrix M , we have*

$$\mathcal{L}_{\text{nc,sm}}(\text{ShiftedTensor}(M)) \geq \frac{\text{PT-rank}(M)}{n^{d-\log d+1}}.$$

4.1 The complexity measure ρ on order $2d$ tensors

We introduce a complexity measure on tensors $A : [n]^{2d} \rightarrow \mathbb{F}$, denoted $\rho(A)$, which is closely related to PT-rank (normalized by factor n^{-d}), but with an added “max” quantifier that affects the \top_1 and \top_k partial transposes. (A precise relationship between ρ and PT-rank is noted at the end of this section.)

► **Definition 4.3.** *For any $a, b \in \{0, 1\}$ and $\alpha \in \{0, 1\}^{d-1}$, we define a partition $I_{a,\alpha,b} \sqcup J_{a,\alpha,b} = [2d]$ by*

$$\begin{aligned} I_{a,\alpha,b} &:= [2d] \cap \{ a, 2 + \alpha_1, 4 + \alpha_2, \dots, 2d - 2 + \alpha_d, 2d + b \}, \\ J_{a,\alpha,b} &:= [2d] \cap \{ 1 - a, 3 - \alpha_1, 5 - \alpha_2, \dots, 2d - 1 - \alpha_d, 2d + 1 - b \}. \end{aligned}$$

For a tensor $A : [n]^{2d} \rightarrow \mathbb{F}$, we define its relative rank with respect to $(a, \alpha, b) \in \{0, 1\}^{d+1}$ as follows:

$$\text{relrk}_{a,\alpha,b}(A) := n^{-d} \text{rank}(\text{Mat}_{I_{a,\alpha,b}, J_{a,\alpha,b}}(A)).$$

► **Lemma 4.4.** $\text{relrk}_{a,\alpha,b}(A) \leq n^{-\mathbb{1}[a \neq b]}$.

Proof. Follows from the observation that

$$\begin{aligned} |I_{a,\alpha,b}| &= d - 1 + \mathbb{1}[a = 1] + \mathbb{1}[b = 0], \\ |J_{a,\alpha,b}| &= d - 1 + \mathbb{1}[a = 0] + \mathbb{1}[b = 1], \end{aligned}$$

and therefore $\min\{|I_{a,\alpha,b}|, |J_{a,\alpha,b}|\} = d - \mathbb{1}[a \neq b]$. ◀

The next lemmas show that relrk is sub-additive and multiplicative.

► **Lemma 4.5.** $\text{relrk}_{a,\alpha,b}(\sum_i A_i) \leq \sum_i \text{relrk}_{a,\alpha,b}(A_i)$.

Proof. Follows from sub-additivity of rank. ◀

113:12 Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $\text{VNC}^1 \neq \text{VNP}$

► **Lemma 4.6.** For all $A : [n]^{2d} \rightarrow \mathbb{F}$ and $B : [n]^{2e} \rightarrow \mathbb{F}$ and $a, b, c \in \{0, 1\}$ and $\alpha \in \{0, 1\}^{d-1}$ and $\beta \in \{0, 1\}^{e-1}$, with respect to the tensor product $A \otimes B : [n]^{2(d+e)} \rightarrow \mathbb{F}$, we have

$$\begin{aligned} I_{a,\alpha,b,\beta,c} &= I_{a,\alpha,b} \sqcup I_{b,\beta,c}, \\ J_{a,\alpha,b,\beta,c} &= J_{a,\alpha,b} \sqcup J_{b,\beta,c}, \\ \text{Mat}_{I_{a,\alpha,b,\beta,c}, J_{a,\alpha,b,\beta,c}}(A \otimes B) &= \text{Mat}_{I_{a,\alpha,b}, J_{a,\alpha,b}}(A) \boxtimes \text{Mat}_{I_{b,\beta,c}, J_{b,\beta,c}}(B), \\ \text{relrk}_{a,\alpha,b,\beta,c}(A \otimes B) &= \text{relrk}_{a,\alpha,b}(A) \text{relrk}_{b,\beta,c}(B). \end{aligned}$$

(Note: We only require the inequality $\text{relrk}_{a,\alpha,b,\beta,c}(A \otimes B) \leq \text{relrk}_{a,\alpha,b}(A) \text{relrk}_{b,\beta,c}(B)$, that is, the sub-multiplicativity of relrk .)

Proof. Follows from definitions and multiplicativity of matrix rank under Kronecker product. ◀

► **Definition 4.7.** For a tensor $A : [n]^{2d} \rightarrow \mathbb{F}$, notation $\{X_\alpha\} \vdash A$ denotes that X_α are tensors $[n]^{2d} \rightarrow \mathbb{F}$, indexed by $\alpha \in \{0, 1\}^{d-1}$, such that $A = \sum_\alpha X_\alpha$.

We define a complexity measure $\rho(A)$ by

$$\rho(A) := \max_{a,b \in \{0,1\}} \min_{\{X_\alpha\} \vdash A} \sum_{\alpha \in \{0,1\}^{d-1}} \text{relrk}_{a,\alpha,b}(X_\alpha).$$

► **Lemma 4.8.** ρ is sub-additive, that is, $\rho(\sum_i A_i) \leq \sum_i \rho(A_i)$ where $\{A_i\}_i$ is any indexed family of tensors $A_i : [n]^{2d} \rightarrow \mathbb{F}$.

Proof. We have

$$\begin{aligned} \rho\left(\sum_i A_i\right) &= \max_{a,b} \min_{\{X_\alpha\} \vdash \sum_i A_i} \sum_\alpha \text{relrk}_{a,\alpha,b}(X_\alpha) \\ &\leq \max_{a,b} \min_{\substack{\{X_{i,\alpha}\} \vdash A_i \\ X_\alpha := \sum_i X_{i,\alpha}}} \sum_\alpha \text{relrk}_{a,\alpha,b}(X_\alpha) \\ &= \max_{a,b} \min_{\{X_{i,\alpha}\} \vdash A_i} \sum_\alpha \text{relrk}_{a,\alpha,b}\left(\sum_i X_{i,\alpha}\right) \\ &\leq \max_{a,b} \min_{\{X_{i,\alpha}\} \vdash A_i} \sum_i \sum_\alpha \text{relrk}_{a,\alpha,b}(X_{i,\alpha}) \\ &= \max_{a,b} \sum_i \min_{\{X_{i,\alpha}\} \vdash A_i} \sum_\alpha \text{relrk}_{a,\alpha,b}(X_{i,\alpha}) \\ &\leq \sum_i \max_{a,b} \min_{\{X_{i,\alpha}\} \vdash A_i} \sum_\alpha \text{relrk}_{a,\alpha,b}(X_{i,\alpha}) \\ &= \sum_i \rho(A_i). \end{aligned}$$

Unlike relative rank, ρ is not obviously multiplicative or sub-multiplicative under tensor product. However, the following key lemma shows that ρ *always* shrinks under tensor products. In contrast, the relative rank of a tensor product $A \otimes B$ is only guaranteed to shrink when (the matrix flattening of) A or B has unbalanced dimensions.

► **Lemma 4.9.** For all $A : [n]^{2d} \rightarrow \mathbb{F}$ and $B : [n]^{2e} \rightarrow \mathbb{F}$, we have

$$\rho(A \otimes B) \leq n^{-1} \min\{\rho(A), \rho(B)\}.$$

Proof. Let a, b, c range over $\{0, 1\}$, and let α, β range over $\{0, 1\}^{d-1}, \{0, 1\}^{e-1}$ respectively. We have

$$\begin{aligned}
\rho(A \otimes B) &= \max_{a,c} \min_{\{Z_{\alpha,b,\beta}\} \vdash A \otimes B} \sum_{\alpha,b,\beta} \text{relrk}_{a,\alpha,b,\beta,c}(Z_{\alpha,b,\beta}) \\
&\leq \max_{a,c} \min_{\substack{\{X_\alpha\} \vdash A, \{Y_\beta\} \vdash B \\ Z_{\alpha,b,\beta} := \mathbf{1}[b=1-c] \cdot X_\alpha \otimes Y_\beta}} \sum_{\alpha,b,\beta} \text{relrk}_{a,\alpha,b,\beta,c}(Z_{\alpha,b,\beta}) \\
&= \max_{a,c} \min_{\{X_\alpha\} \vdash A, \{Y_\beta\} \vdash B} \sum_{\alpha,\beta} \text{relrk}_{a,\alpha,1-c,\beta,c}(X_\alpha \otimes Y_\beta) \\
&= \max_{a,c} \min_{\{X_\alpha\} \vdash A, \{Y_\beta\} \vdash B} \sum_{\alpha,\beta} \text{relrk}_{a,\alpha,1-c}(X_\alpha) \text{relrk}_{1-c,\beta,c}(Y_\beta) \quad (\text{Lemma 4.6}) \\
&\leq \max_{a,c} \min_{\{X_\alpha\} \vdash A, \{Y_\beta\} \vdash B} \sum_{\alpha,\beta} \text{relrk}_{a,\alpha,1-c}(X_\alpha) n^{-1} \quad (\text{Lemma 4.4}) \\
&= n^{-1} \max_{a,c} \min_{\{X_\alpha\} \vdash A} \sum_{\alpha} \text{relrk}_{a,\alpha,1-c}(X_\alpha) \\
&= n^{-1} \max_{a,b} \min_{\{X_\alpha\} \vdash A} \sum_{\alpha} \text{relrk}_{a,\alpha,b}(X_\alpha) \\
&= n^{-1} \rho(A). \quad \blacktriangleleft
\end{aligned}$$

The proof of $\rho(A \otimes B) \leq n^{-1} \rho(B)$ follows similarly.

4.2 Non-commutative formula lower bounds from ρ

► **Definition 4.10** (The order d flattening $A^b : [n^2]^d \rightarrow \mathbb{F}$ of an order $2d$ tensor $A : [n]^{2d} \rightarrow \mathbb{F}$). For an order $2d$ tensor $A : [n]^{2d} \rightarrow \mathbb{F}$, we define its order d flattening $A^b : [n^2]^d \rightarrow \mathbb{F}$ by the identity

$$A(a_1, b_1, \dots, a_d, b_d) = A^b(\langle a_1, b_1 \rangle, \dots, \langle a_d, b_d \rangle)$$

for all $a_1, b_1, \dots, a_d, b_d \in [n]$.

Note that this flattening operation commutes with the tensor product: for any $A : [n]^{2d} \rightarrow \mathbb{F}$ and $B : [n]^{2e} \rightarrow \mathbb{F}$, we have $(A \otimes B)^b = A^b \otimes B^b$ (as tensors $[n]^{2(d+e)} \rightarrow \mathbb{F}$).

► **Theorem 4.11.** For every $A : [n]^{2d} \rightarrow \mathbb{F}$, we have $\mathcal{L}_{\text{nc,sm}}(A^b) \geq n^{\log d} \rho(A)$. That is, any non-commutative set-multilinear formula computing the flattened tensor $A^b : [n^2]^d \rightarrow \mathbb{F}$ has size at least $n^{\log d} \rho(A)$.

Proof. We argue by induction on d . The base case $d = 1$ is trivial. Assume $d \geq 2$ and fix $\{(e_i, f_i, B_i, C_i)\}_i$ with $e_i, f_i \geq 1$ and $e_i + f_i = d$ and $B_i : [n]^{e_i} \rightarrow \mathbb{F}$ and $C_i : [n]^{f_i} \rightarrow \mathbb{F}$ such that $A = \sum_i B_i \otimes C_i$ and $\mathcal{L}_{\text{nc,sm}}(A^b) = \sum_i (\mathcal{L}_{\text{nc,sm}}(B_i^b) + \mathcal{L}_{\text{nc,sm}}(C_i^b))$.

We have

$$\begin{aligned}
\mathcal{L}_{\text{nc,sm}}(A^b) &\geq \sum_i \left(n^{\log e_i} \rho(B_i) + n^{\log f_i} \rho(C_i) \right) \quad (\text{induction hypothesis}) \\
&\geq \sum_{i: e_i \geq d/2} n^{\log e_i} \rho(B_i) + \sum_{i: f_i > d/2} n^{\log f_i} \rho(C_i) \\
&\geq \sum_{i: e_i \geq d/2} n^{\log e_i + 1} \rho(B_i \otimes C_i) + \sum_{i: f_i > d/2} n^{\log f_i + 1} \rho(B_i \otimes C_i) \quad (\text{Lemma 4.9})
\end{aligned}$$

113:14 Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $\text{VNC}^1 \neq \text{VNP}$

$$\begin{aligned}
&\geq n^{\log d} \sum_i \rho(B_i \otimes C_i) \\
&\geq n^{\log d} \rho\left(\sum_i B_i \otimes C_i\right) \quad (\text{Lemma 4.8}) \\
&= n^{\log d} \rho(A). \quad \blacktriangleleft
\end{aligned}$$

We are finally ready to prove Theorem 4.2.

Proof of Theorem 4.2. Let M be an $n^d \times n^d$ matrix. Recall the order $d + 1$ tensor $\text{ShiftedTensor}(M) : [n^2]^{d+1} \rightarrow \mathbb{F}$, which equals $\text{PaddedTensor}(M)^b$ for the order $2d + 2$ tensor $\text{PaddedTensor}(M) : [n]^{2d+2} \rightarrow \mathbb{F}$ defined by

$$\begin{aligned}
\text{PaddedTensor}(M)(p, i_1, j_1, i_2, \dots, j_{d-1}, i_d, j_d, q) &= \mathbf{1}\{p = q = 1\} \cdot M_{(i_1, \dots, i_d), (j_1, \dots, j_d)}, \\
\text{ShiftedTensor}(M)(\langle p, i_1 \rangle, \langle j_1, i_2 \rangle, \dots, \langle j_{d-1}, i_d \rangle, \langle j_d, q \rangle) &= \mathbf{1}\{p = q = 1\} \cdot M_{(i_1, \dots, i_d), (j_1, \dots, j_d)}.
\end{aligned}$$

It follows immediately from definitions of PT-rank and ρ that $\text{PT-rank}(M) = n^{d+1} \rho(\text{PaddedTensor}(M))$. Theorem 4.11 now implies the desired bound

$$\mathcal{L}_{\text{nc,sm}}(\text{ShiftedTensor}(M)) \geq n^{\log d} \rho(\text{PaddedTensor}(M)) = \frac{\text{PT-rank}(M)}{n^{d-\log d+1}}. \quad \blacktriangleleft$$

5 Commutative set-multilinear formula lower bounds from PT-rank

This section generalizes the argument in the last section to prove our main theorem (Theorem 1.5).

5.1 Extending ρ to tensors over subgraphs of Path_d

Instead of considering tensors $[n]^{2d} \rightarrow \mathbb{F}$, we consider tensors $[n]^{D(G)} \rightarrow \mathbb{F}$ with coordinates indexed by the (symmetric) directed edge set of a subgraph G of an undirected path graph.

Let $\text{Path}_{\mathbb{Z}}$ be the undirected path graph with vertex set $\{v_i : i \in \mathbb{Z}\}$ and edge set $\{\{v_{i-1}, v_i\} : i \in \mathbb{Z}\}$.

Let $G = (V(G), E(G))$, $E(G) \subseteq \binom{V(G)}{2}$, range over nonempty finite subgraphs of $\text{Path}_{\mathbb{Z}}$ with no isolated vertices (i.e., $V(G) = \bigcup_{e \in E(G)} e$). We next introduce some useful notation:

- We consider the partition $V_1(G) \sqcup V_2(G) = V(G)$ defined by

$$\begin{aligned}
V_1(G) &:= \{v \in V(G) : v \text{ has degree 1 in } G\}, \\
V_2(G) &:= \{v \in V(G) : v \text{ has degree 2 in } G\}.
\end{aligned}$$

- Let $D(G) \subseteq V(G) \times V(G)$ be the (symmetric) set of directed edges of G , that is,

$$D(G) := \{vw : \{v, w\} \in E(G)\}.$$

We consider the partition $D_1(G) \sqcup D_2(G) = D(G)$ defined by

$$\begin{aligned}
D_1(G) &:= \{vw \in D(G) : v \in V_1(G)\}, \\
D_2(G) &:= \{vw \in D(G) : v \in V_2(G)\}.
\end{aligned}$$

Note that $|D_1(G)| = |V_1(G)|$ and $|D_2(G)| = 2|V_2(G)|$.

- For any set $S \subseteq V(\text{Path}_{\mathbb{Z}})$, let $\mathcal{P}(S)$ denote the set of pairs $\gamma = (I_\gamma, J_\gamma)$ of disjoint sets $I_\gamma, J_\gamma \subseteq D(\text{Path}_{\mathbb{Z}})$ such that

$$S = \{v : vu \in I_\gamma\} = \{v : vw \in J_\gamma\}.$$

(That is, for each $v_i \in S$ with neighbors v_{i-1}, v_{i+1} in $\text{Path}_{\mathbb{Z}}$, exactly one of the sets I_γ and J_γ contains $v_i v_{i-1}$ and the other contains $v_i v_{i+1}$.) Note that $|I_\gamma| = |J_\gamma| = |S|$ and $|\mathcal{P}(S)| = 2^{|S|}$.

- For disjoint sets $S_1, S_2 \subseteq V(\text{Path}_{\mathbb{Z}})$ and $\gamma_1 \in \mathcal{P}(S_1)$ and $\gamma_2 \in \mathcal{P}(S_2)$, let

$$\gamma_1 \sqcup \gamma_2 := (I_{\gamma_1} \sqcup I_{\gamma_2}, J_{\gamma_1} \sqcup J_{\gamma_2}) \in \mathcal{P}(S_1 \sqcup S_2).$$

We now consider tensors $A : [n]^{D(G)} \rightarrow \mathbb{F}$ and define analogues of relrk and ρ .

► **Remark 5.1.** When G is a path of length d with vertices v_0, \dots, v_d , we identify the sequence of directed edges $v_0 v_1, v_1 v_0, v_1 v_2, v_2 v_1, \dots, v_{d-1} v_d, v_d v_{d-1}$ with integers $1, \dots, 2d$. We thus identify $D(G)$ with the set $[2d]$ and tensors $[n]^{D(G)} \rightarrow \mathbb{F}$ with tensors $[n]^{2d} \rightarrow \mathbb{F}$. Under this identification, the following definitions of relrk and ρ directly generalize those in §4.

► **Definition 5.2.** For a graph $G \subseteq \text{Path}_{\mathbb{Z}}$ and tensor $A : [n]^{D(G)} \rightarrow \mathbb{F}$ and $\alpha \in \mathcal{P}(V_2(G))$ and $\gamma \in \mathcal{P}(V_1(G))$, we define

$$\text{relrk}_{\alpha, \gamma}(A) := n^{-|E(G)|} \text{rank}(\text{Mat}_{I_\alpha \sqcup (I_\gamma \cap D_1(G)), J_\alpha \sqcup (J_\gamma \cap D_1(G))}(A)).$$

► **Definition 5.3.** For a graph $G \subseteq \text{Path}_{\mathbb{Z}}$ and tensor $A : [n]^{D(G)} \rightarrow \mathbb{F}$, notation $\{X_\alpha\} \vdash A$ denotes that X_α are tensors $[n]^{D(G)} \rightarrow \mathbb{F}$, indexed by $\alpha \in \mathcal{P}(V_2(G))$, such that $A = \sum_\alpha X_\alpha$.

We define the complexity measure $\rho(A)$ by

$$\rho(A) := \max_{\gamma \in \mathcal{P}(V_1(G))} \min_{\{X_\alpha\} \vdash A} \sum_{\alpha \in \mathcal{P}(V_2(G))} \text{relrk}_{\alpha, \gamma}(X_\alpha).$$

► **Lemma 5.4.** $\rho(\sum_i A_i) \leq \sum_i \rho(A_i)$.

Note that $|I_\alpha| = |J_\alpha| = |V_2(G)|$ and $|D_1(G) \cap I_\gamma| + |D_1(G) \cap J_\gamma| = |V_1(G)|$. As a consequence, we get the following generalization of Lemma 4.4.

► **Lemma 5.5.** $\text{relrk}_{\alpha, \gamma}(A) \leq n^{-||I_\gamma \cap D_1(G)| - |J_\gamma \cap D_1(G)||}$.

Part (a) of the next lemma shows that the bound of Lemma 5.5 may be as small as $n^{-|V_1(G)|}$. Part (b) makes a related observation that will be useful later.

► **Lemma 5.6.**

(a) For every graph $G \subseteq \text{Path}_{\mathbb{Z}}$, there exist unique elements $\gamma^+, \gamma^- \in \mathcal{P}(V_1(G))$ such that

$$I_{\gamma^+} = J_{\gamma^-} = D_1(G), \quad J_{\gamma^+} \cap D_1(G) = I_{\gamma^-} \cap D_1(G) = \emptyset.$$

(b) For all edge-disjoint graphs $G, H \subseteq \text{Path}_{\mathbb{Z}}$, there exist unique elements $\delta^+, \delta^- \in \mathcal{P}(V_1(G) \cap V_1(H))$ such that

$$I_{\delta^+} = J_{\delta^-} = D_1(G) \cap D_1(H), \quad J_{\delta^+} \cap D_1(H) = I_{\delta^-} \cap D_1(H) = \emptyset.$$

Proof.

(a) Consider any $v \in V_1(G)$ with neighbors u, w in $\text{Path}_{\mathbb{Z}}$. Without loss of generality, $\{v, u\} \notin E(G)$ and $\{v, w\} \in E(G)$. We include vw in I_{γ^+} ($= J_{\gamma^-}$) and vu in I_{γ^-} ($= J_{\gamma^+}$).

113:16 Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $\text{VNC}^1 \neq \text{VNP}$

- (b) Consider any $v \in V_1(G) \cap V_1(H)$ with neighbors u, w in $\text{Path}_{\mathbb{Z}}$. Without loss of generality, $\{v, u\} \in E(G) \setminus E(H)$ and $\{v, w\} \in E(H) \setminus E(G)$. We include vw in I_{δ^+} ($= J_{\delta^-}$) and vu in I_{δ^-} ($= J_{\delta^+}$). \blacktriangleleft

► **Lemma 5.7.** *Suppose that G, H are edge-disjoint finite subgraphs of $\text{Path}_{\mathbb{Z}}$, let $A : [n]^{D(G)} \rightarrow \mathbb{F}$ and $B : [n]^{D(H)} \rightarrow \mathbb{F}$, and consider the tensor product $A \otimes B : [n]^{D(G \cup H)} \rightarrow \mathbb{F}$. For all*

$$\begin{aligned} \alpha &\in \mathcal{P}(V_2(G)), & \beta &\in \mathcal{P}(V_2(H)), & \delta &\in \mathcal{P}(V_1(G) \cap V_1(H)), \\ \xi &\in \mathcal{P}(V_1(G) \setminus V_1(H)), & \zeta &\in \mathcal{P}(V_1(H) \setminus V_1(G)), \end{aligned}$$

we have

$$\text{relrk}_{\alpha \cup \beta \cup \delta, \xi \cup \zeta}(A \otimes B) = \text{relrk}_{\alpha, \xi \cup \delta}(A) \text{relrk}_{\beta, \zeta \cup \delta}(B).$$

► **Lemma 5.8.** *For all edge-disjoint $G, H \subseteq \text{Path}_{\mathbb{Z}}$ and tensors $A : [n]^{D(G)} \rightarrow \mathbb{F}$ and $B : [n]^{D(H)} \rightarrow \mathbb{F}$, we have*

$$\rho(A \otimes B) \leq n^{-|V_1(G) \cap V_1(H)|} \min\{\rho(A), \rho(B)\}.$$

(See full paper for the proof, which is similar to Lemma 5.7.)

5.2 Set-multilinear formula lower bounds from ρ

The next definition generalizes the flattening operation \flat introduced in Def. 4.10.

► **Definition 5.9.** *For any $G \subseteq \text{Path}_{\mathbb{Z}}$ and $A : [n]^{D(G)} \rightarrow \mathbb{F}$, let $A^\flat : [n^2]^{E(G)} \rightarrow \mathbb{F}$ denote the flattened tensor defined in the natural way: for $x \in [n]^{D(G)}$, we have $A^\flat(y) = A(x)$ where $y \in [n^2]^{E(G)}$ is the element defined by $y_{\{v_{i-1}, v_i\}} = \langle x_{v_{i-1}}, x_{v_i} \rangle$ for each $i \in \mathbb{Z}$ such that $\{v_{i-1}, v_i\} \in E(G)$.*

The analog of Theorem 4.11 for (commutative) set-multilinear formulas has a similar proof based on the inductive definition of \mathcal{L}_{sm} . However, instead of $\log d$ in the exponent of n , the relevant parameter becomes $\log \ell(G)$ where $\ell(G)$ is the length of the longest path in G . We record this notation below:

► **Definition 5.10.** *For $G \subseteq \text{Path}_{\mathbb{Z}}$, let $\ell(G)$ denote the length of the longest path in G (i.e., the maximal number of edges in a connected component of G).*

► **Theorem 5.11.** *For every $G \subseteq \text{Path}_{\mathbb{Z}}$ and $A : [n]^{D(G)} \rightarrow \mathbb{F}$, we have*

$$\mathcal{L}_{\text{sm}}(A^\flat) \geq n^{\log \ell(G)} \rho(A).$$

That is, any set-multilinear arithmetic formula computing the flattened tensor $A^\flat : [n^2]^{E(G)} \rightarrow \mathbb{F}$ has size at least $n^{1+\log \ell(G)} \rho(A)$.

Proof. We argue by induction on $|E(G)|$. The base case $|E(G)| = 1$ is trivial, since $|E(G)| - |V(G)| + 1 + \log |E(G)| = 0$ and $\rho(A) \leq 1$.

Now assume $|E(G)| \geq 2$ and fix $\{(H_i, K_i, B_i, C_i)\}_i$ with $H_i, K_i \subsetneq G$ such that $H_i \cup K_i = G$ and $B_i : [n]^{D(H_i)} \rightarrow \mathbb{F}$ and $C_i : [n]^{D(K_i)} \rightarrow \mathbb{F}$ such that $A = \bigcup_i B_i \otimes C_i$ and $\mathcal{L}_{\text{sm}}(A^\flat) = \sum_i (\mathcal{L}_{\text{sm}}(B_i^\flat) + \mathcal{L}_{\text{sm}}(C_i^\flat))$.

Note that for all i ,

$$\ell(G) \leq \max\{\ell(H_i), \ell(K_i)\} \cdot (|V_1(H_i) \cap V_1(K_i)| + 1).$$

Therefore,

$$\log \ell(G) \leq \max\{\log \ell(H_i), \log \ell(K_i)\} + |V_1(H_i) \cap V_1(K_i)|.$$

We now have

$$\begin{aligned} \mathcal{L}_{\text{sm}}(A^b) &\geq \sum_i \left(n^{\log \ell(H_i)} \rho(B_i) + n^{\log \ell(K_i)} \rho(C_i) \right) \quad (\text{induction hypothesis}) \\ &\geq \sum_i n^{\max\{\log \ell(H_i), \log \ell(K_i)\} + |V_1(H_i) \cap V_1(K_i)|} \rho(B_i \otimes C_i) \quad (\text{Lemma 5.8}) \\ &\geq n^{\log \ell(G)} \sum_i \rho(B_i \otimes C_i) \\ &\geq n^{\log \ell(G)} \rho\left(\sum_i B_i \otimes C_i\right) \quad (\text{Lemma 5.4}) \\ &= n^{\log \ell(G)} \rho(A). \quad \blacktriangleleft \end{aligned}$$

In the special case where G is path of length d , Theorem 5.11 has the following corollary, which directly strengthens Theorem 4.11 by replacing $\mathcal{L}_{\text{nc,sm}}$ with \mathcal{L}_{sm} .

► **Corollary 5.12.** *For every $A : [n]^{2d} \rightarrow \mathbb{F}$, we have $\mathcal{L}_{\text{sm}}(A^b) \geq n^{\log d} \rho(A)$.*

Finally, Theorem 1.5 follows directly from Theorem 4.11 via the same observation that

$$\text{PT-rank}(M) = n^{d+1} \rho(\text{PaddedTensor}(M))$$

for every $n^d \times n^d$ matrix M and that $\text{ShiftedTensor}(M) = \text{PaddedTensor}(M)^b$.

6 Upper bounds and candidate hard matrices for PT-rank

In this section, we first present some known PT-rank upper bounds for certain families of matrices. These bounds are based on Hrubeš' recent subquadratic upper bound for the sum-of-squares problem [14]. Then, we give some explicit candidate hard matrices that might realize our desired lower bound.

6.1 Kronecker products of $n \times n$ matrices have low PT-rank

Recall Hrubeš' upper bound $\text{SoS}(q_n) = O(n^{1.62})$ (Theorem 1.1). By Proposition 3.15, this is equivalent to $\text{PT-rank}(I_{n^2}) = O(n^{1.62})$. This immediately gives the following upper bound on $\text{PT-rank}(I_{n^d})$ for all d .

► **Proposition 6.1.** $\text{PT-rank}(I_{n^d}) = n^{0.81d+O(1)}$.

Proof. We prove the bound $\text{PT-rank}(I_{n^d}) = O(n^{0.81d})$ in the case where d is even. (The bound $n^{0.81d+O(1)}$ for odd d follows by monotonicity.) This bound is given by the following calculation:

$$\text{PT-rank}_{[n]^d}(I_{n^d}) \leq \text{PT-rank}_{[n^{d/2}]^2}(I_{(n^{d/2})^2}) = O((n^{d/2})^{1.62}) = O(n^{0.81d}).$$

The first inequality holds since $\text{PT-rank}_{[n]^d}$ can use \top_κ for all $\kappa \subseteq [d-1]$, whereas $\text{PT-rank}_{[n^{d/2}]^2}$ can only use \top_κ for $\kappa \in \{\emptyset, [d/2]\}$. ◀

Notice that the PT-rank bound of Proposition 6.1 involves only two partial transposes, namely \top_\emptyset and $\top_{[d/2]}$. Using all 2^d partial transposes, we can show a much stronger upper bound:

► **Proposition 6.2.** $\text{PT-rank}(I_{n^d}) = n^{O(d^{0.7})}$.

Proof. The construction by Hrubeš shows that

$$\prod_{i=1}^d \left(\sum_{j=1}^n X_{i,j}^2 \right) = \prod_{i'=1}^{d/2} \left(\sum_{j'=1}^{O(n^{1.62})} f_{i',j'}^2 \right),$$

where $f_{i',j'}$ is a bilinear form in $\{X_{i,(2j'-1)}\}_{i=1}^d$ and $\{X_{i,(2j')}\}_{i=1}^d$. We may repeat the step by considering $f_{i',j'}$ as new variables. Inductively, this gives

$$\prod_{i=1}^d \left(\sum_{j=1}^n X_{i,j}^2 \right) = \sum_{j'=1}^s g_{j'}^2,$$

where $g_{j'}$ is multilinear in the d sets of variables and

$$s = O(n^{(1.62)^{\log_2 d}}) \leq n^{O(d^{0.7})}.$$

The corollary follows from Proposition 3.15. ◀

The upper bound above extends to Kronecker products of $n \times n$ matrices, via $GL(n, \mathbb{F})^d$ -invariance of PT-rank (Lemma 3.8). Thus, we have proved:

► **Theorem 6.3.** *If M is a Kronecker product of d $n \times n$ matrices, then $\text{PT-rank}(M)$ is at most $n^{O(d^{0.7})}$.*

6.2 Matrices in VBP_{ord} have low PT-rank

The Kronecker products of $n \times n$ matrices have low PT-rank. Using this fact, we may derive an upper bound for all matrices whose shifted tensors are in the complexity class VBP_{ord} .

► **Definition 6.4** (Ordered algebraic branching programs). *Let f be a set-multilinear polynomial over \mathbb{F} of degree d in variable sets $X^{(1)}, X^{(2)}, \dots, X^{(d)}$. We say that f has an ordered algebraic branching program of width w and size s if*

$$f = v_1^\top M_1 M_2 \cdots M_d v_2,$$

where $v_1 \in \mathbb{F}^{w_0}$ and $v_2 \in \mathbb{F}^{w_d}$ are vectors whose entries are in \mathbb{F} , $M_i \in \mathbb{F}^{w_{i-1} \times w_i}$ are matrices whose entries are linear combinations of variables in X_i for $i \in [d]$, and the following holds:

$$w_i \leq w \text{ for } 0 \leq i \leq [d] \text{ and } \sum_{i=0}^d w_i \leq s.$$

► **Definition 6.5** (VBP_{ord}). *A sequence of set-multilinear polynomials with n variables is in VBP_{ord} if and only if it can be computed by a sequence of ordered algebraic branching programs with size $s = n^{O(1)}$.*

An ordered algebraic branching program for a tensor A is an ordered algebraic branching program for its associated set-multilinear polynomial P_A (as defined in Section 2.3). The class VBP_{ord} is defined for tensors similarly.

Our main result (Theorem 1.5) shows that sufficiently strong lower bounds on $\text{PT-rank}(M)$ for explicit matrices M would imply a separation of classes VNC^1 and VNP . For a more stringent notion of “explicit”, this reduction could even potentially separate VNC^1 from VP (i.e., arithmetic formulas from circuits). In this section, however, we describe a limitation of PT-rank by showing that it cannot separate VNC^1 from VBP_{ord} (i.e., arithmetic formulas from ordered branching programs).

► **Theorem 6.6.** *For any sequence of $n^d \times n^d$ matrices M such that $\text{ShiftedTensor}(M) \in \text{VBP}_{\text{ord}}$, we have $\text{PT-rank}(M) \leq n^{0.81d+O(1)}$.*

(See full paper for the proof.)

6.3 Candidate hard matrices

We now propose a natural family of $n^d \times n^d$ matrices over $\mathbb{F} = \mathbb{C}$ which are plausible candidates for having nearly maximal PT-rank.

► **Definition 6.7.** *Let d be an even integer. Let n be a prime power such that $n > 2d$. Let T be an $d \times d$ matrix over the prime field \mathbb{F}_n . Let $\omega = e^{2\pi i/n}$ be a primitive n th root of unity. Then, define the $n^d \times n^d$ matrix W_T over \mathbb{C} by*

$$(W_T)_{(i_1, \dots, i_d), (j_1, \dots, j_d)} = \omega^{(i_1, \dots, i_d)T(j_1, \dots, j_d)^\top}.$$

With the definition above, we first consider three choices T_1, T_2, T_3 that don't work, that is, where W_T has small PT-rank.

Identity matrix

For $T_1 = I_d$, we have

$$(W_{T_1})_{(i_1, \dots, i_d), (j_1, \dots, j_d)} = \omega^{(i_1, \dots, i_d)(j_1, \dots, j_d)^\top}.$$

We may define $n \times n$ matrices $W_{T_1, l}$ for $l \in [d]$ by $(W_{T_1, l})_{i_i, j_i} = \omega^{i_i j_i}$. Then, we have $W_{T_1} = W_{T_1, 1} \boxtimes \dots \boxtimes W_{T_1, d}$. By Theorem 6.3, this implies $\text{PT-rank}(W_{T_1}) = n^{O(d^{0.7})}$.

Cyclic matrix

Consider the cyclic permutation matrix T_2 defined by

$$(T_2)_{a,b} = 1[a + 1 \equiv b \pmod{d}].$$

For this choice of T_2 , we have

$$(W_{T_2})_{(i_1, \dots, i_d), (j_1, \dots, j_d)} = \omega^{(i_1, \dots, i_{d-1}, i_d)(j_2, \dots, j_d, j_1)^\top}.$$

Let $\kappa = \{1, 3, 5, \dots, d-1\}$. Then, by definition of partial transpose, we have

$$(W_{T_2}^\top)_{(j_1, i_2, \dots, j_{d-1}, i_d), (i_1, j_2, \dots, i_{d-1}, j_d)} = (W_{T_2})_{(i_1, \dots, i_d), (j_1, \dots, j_d)} = \omega^{(i_1, \dots, i_{d-1}, i_d)(j_2, \dots, j_d, j_1)^\top}.$$

Define n^d vectors u and v by

$$u_{(j_1, i_2, \dots, j_{d-1})} = \omega^{(i_2, i_4, \dots, i_d)(j_1, j_3, \dots, j_{d-1})^\top}, v_{(i_1, j_2, \dots, i_{d-1})} = \omega^{(i_1, i_3, \dots, i_{d-1})(j_2, j_4, \dots, j_d)^\top}.$$

We then have $W_{T_2}^\top = uv^\top$, which means it has rank 1. Therefore, $\text{PT-rank}(W_{T_2}) = 1$.

Upper triangular matrix

Consider the upper triangular matrix T_3 defined by

$$(T_3)_{a,b} = 1[a \leq b].$$

For this choice of T_3 , we have $(W_{T_3})_{(i_1, \dots, i_d), (j_1, \dots, j_d)} = \omega^{\sum_{k \leq l} i_k j_l}$.

Let A be the unique $[n]^{2d}$ tensor such that $A^b = \text{ShiftedTensor}(W_{T_3})$. Then,

$$A(p, i_1, j_1, i_2, \dots, j_d, q) = (W_{T_3})_{(i_1, \dots, i_d), (j_1, \dots, j_d)}.$$

Then, it can be observed that $\text{rank}(\text{Mat}_{\{1, \dots, d+1\}, \{d+2, \dots, 2d+2\}}) \leq n^2$. (In fact, $A^b \in \text{VBP}_{\text{ord}}$.) Therefore, $\text{PT-rank}(W_{T_3}) \leq n^{0.81d+O(1)}$ by Theorem 6.6.

The three matrices T_1, T_2, T_3 considered above give rise to W_T with small PT-rank. We now give some evidence that a *Cauchy matrix* (in which every submatrix has full rank) might be a good choice of T . (See full paper for proofs of the propositions in this section.)

► **Definition 6.8.** For $\lambda \subseteq [d]$, let $W_T^{[\lambda]}$ be the $n^{2|\lambda|} \times n^{2(d-|\lambda|)}$ matrix defined by

$$(W_T^{[\lambda]})_{(i_\lambda, j_\lambda), (i_{\lambda^c}, j_{\lambda^c})} = (W_T)_{(i_1, \dots, i_d), (j_1, \dots, j_d)}.$$

► **Proposition 6.9.** Let T be a Cauchy matrix. For every $\lambda \subseteq [d]$, the matrix $W_T^{[\lambda]} \in \mathbb{C}^{n^{2|\lambda|} \times n^{2(d-|\lambda|)}}$ has full rank:

$$\text{rank}(W_T^{[\lambda]}) = n^{2 \min\{|\lambda|, d-|\lambda|\}}.$$

The two propositions above disable the only two upper bound techniques we know: taking a single partial transpose, and finding some flattening of K_T with small rank in order to apply Hrubeš' construction.

References

- 1 John F Adams and Michael F Atiyah. K-theory and the hopf invariant. *The Quarterly Journal of Mathematics*, 17(1):31–38, 1966.
- 2 Prashanth Amireddy, Ankit Garg, Neeraj Kayal, Chandan Saha, and Bhargav Thankey. Low-Depth Arithmetic Circuit Lower Bounds: Bypassing Set-Multilinearization. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ICALP.2023.12.
- 3 V Arvind and S Raja. Some lower bound results for set-multilinear arithmetic computations. *Chicago Journal OF Theoretical Computer Science*, 6:1–262, 2016.
- 4 Peter Bürgisser, Michael Clausen, and Mohammad A Shokrollahi. *Algebraic complexity theory*, volume 315. Springer Science & Business Media, 2013.
- 5 Marco L. Carmosino, Russell Impagliazzo, Shachar Lovett, and Ivan Mihajlin. Hardness Amplification for Non-Commutative Arithmetic Circuits. In Rocco A. Servedio, editor, *33rd Computational Complexity Conference (CCC 2018)*, volume 102 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 12:1–12:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.CCC.2018.12.
- 6 Prerona Chatterjee. Separating abps and some structured formulas in the non-commutative setting. In *36th Computational Complexity Conference*, 2021.
- 7 Daniel Dugger and Daniel C Isaksen. The hopf condition for bilinear forms over arbitrary fields. *Annals of mathematics*, pages 943–964, 2007.
- 8 Pranjal Dutta, Nitin Saxena, and Thomas Thierauf. Weighted sum-of-squares lower bounds for univariate polynomials imply $\text{VP} \neq \text{VNP}$. *computational complexity*, 33(1):1–54, 2024.
- 9 Zeev Dvir, G Malod, S Perifel, and A Yehudayoff. Separating multilinear branching programs and formulas. In *44th Annual ACM Symposium on Theory of Computing, STOC'12*, 2012.

- 10 Klim Efremenko, Ankit Garg, Rafael Oliveira, and Avi Wigderson. Barriers for Rank Methods in Arithmetic Complexity. In Anna R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ITCS.2018.1.
- 11 Ankit Garg, Visu Makam, Rafael Oliveira, and Avi Wigderson. More barriers for rank methods, via a “numeric to symbolic” transfer. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 824–844. IEEE Computer Society, 2019.
- 12 Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. On the necessary and sufficient conditions for separability of mixed quantum states. *Phys. Lett. A*, 223(1), 1996.
- 13 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.
- 14 Pavel Hrubeš. A subquadratic upper bound on sum-of-squares composition formulas. In *39th Computational Complexity Conference (CCC 2024)*, volume 300 of *LIPIcs*, pages 12:1–12:11, 2024. doi:10.4230/LIPIcs.CCC.2024.12.
- 15 Adolf Hurwitz. Über die komposition der quadratischen formen. *Mathematische Annalen*, 88(1):1–25, 1922.
- 16 Deepanshu Kush and Shubhangi Saraf. Near-Optimal Set-Multilinear Formula Lower Bounds. In Amnon Ta-Shma, editor, *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:33. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.CCC.2023.15.
- 17 Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. On the Partial Derivative Method Applied to Lopsided Set-Multilinear Polynomials. In Shachar Lovett, editor, *37th Computational Complexity Conference (CCC 2022)*, volume 234 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CCC.2022.32.
- 18 Nutan Limaye, Srikanth Srinivasan, and Sebastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 804–814. IEEE Computer Society, 2022.
- 19 Noam Nisan. Lower bounds for non-commutative computation. In *Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 410–418, 1991.
- 20 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Computational complexity*, 6:217–234, 1996. doi:10.1007/BF01294256.
- 21 Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413, 1996.
- 22 AR Rajwade. *Squares*, volume 171. Cambridge University Press, 1993.
- 23 Ran Raz. Multilinear-NC₁ ≠ multilinear-NC₂. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 344–351. IEEE, 2004.
- 24 Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *Journal of the ACM (JACM)*, 60(6):1–15, 2013. doi:10.1145/2535928.
- 25 Ran Raz and Amir Yehudayoff. Balancing syntactically multilinear arithmetic circuits. *Computational Complexity*, 17:515–535, 2008. doi:10.1007/S00037-008-0254-0.
- 26 Ramprasad Satharishi. A survey of lower bounds in arithmetic circuit complexity. *Github survey*, 95, 2015.
- 27 Gadiel Seroussi and Abraham Lempel. Factorization of symmetric matrices and trace-orthogonal bases in finite fields. *SIAM Journal on Computing*, 9(4):758–767, 1980. doi:10.1137/0209059.
- 28 Daniel B Shapiro. *Compositions of quadratic forms*, volume 33. Walter de Gruyter, 2011.
- 29 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. doi:10.1561/04000000039.

113:22 Multi-Quadratic Sum-Of-Squares Lower Bounds Imply $VNC^1 \neq VNP$

- 30 Sébastien Tavenas, Nutan Limaye, and Srikanth Srinivasan. Set-multilinear and non-commutative formula lower bounds for iterated matrix multiplication. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 416–425, 2022. doi:10.1145/3519935.3520044.
- 31 Leslie G Valiant. Completeness classes in algebra. In *Proceedings of the 17th ACM Symposium on Theory of Computing*, pages 249–261, 1979. doi:10.1145/800135.804419.