Electronic Colloquium on Computational Complexity, Report No. 137 (2022)

ON BLOCKY RANKS OF MATRICES

DANIEL AVRAHAM AND AMIR YEHUDAYOFF

Abstract. A matrix is blocky if it is a "blowup" of a permutation matrix. The *blocky rank* of a matrix M is the minimum number of blocky matrices that linearly span M. Hambardzumyan, Hatami and Hatami defined blocky rank and showed that it is connected to communication complexity and operator theory. We describe additional connections to circuit complexity and combinatorics, and we prove upper and lower bounds on blocky rank in various contexts.

1. Introduction

Matrices serve as a model for many objects; linear operators in algebra, communication problems in computational complexity, concept classes in machine learning, and more. There are many ways to measure the complexity of matrices; there are various notions of rank (the "usual" rank, approximate rank, non-negative rank, sign rank, etc.), there are various notions of communication complexity (deterministic, randomized, quantum, etc.), there are various notions in learning theory (VC dimension, Littlestone dimension, margin complexity, etc.), and more. We focus on the notion of *blocky rank* recently defined by Hambardzumyan, Hatami and Hatami [10].

A standard mechanism for defining a complexity measure has two stages. In the first stage, we define the *building blocks* of the model (in our case, matrices of blocky rank one). In the second stage, complexity is defined as the *minimum number of operations* that are needed to generate the target (in our case, sum operations).

Definition. *All identity matrices have blocky rank one. The set of matrices of blocky rank one is also closed under three operations: duplicating a row or a column, permuting the rows or columns, and adding a zero row or a zero column. In other word, a matrix has blocky rank one if up to a permutation of the rows and columns it has blocks of ones of different sizes on the "diagonal" followed by some amount of zeros.*

Definition. *The blocky rank* $\mathsf{blocky}(M)$ *of a matrix* M *is the minimum integer* R *so that* M *can be written as a linear combination of* R *matrices* $B_1, \dots, B_R$*, each of blocky rank one. In this work, we always work over the field* $\mathbb{R}$*.*

Motivation to study blocky rank, and its relatives, comes from various areas. In communication complexity, it is related to understanding randomized communication problems [10]. In operator theory, it is related to idempotents in Schur algebras (see [10] and references within). In circuit complexity, it is related to depth-two threshold circuits. In combinatorics, it is related to covering problems in graphs. In machine learning, it is related to closure properties of Littlestone classes.

ISSN 1433-8092

1.1. **Generic matrices.** A typical first question about complexity is "what is the complexity of a random object?" The "obvious" upper bound on the blocky rank of an $n \times n$ boolean matrix is n, because a boolean matrix with one non-zero row has blocky rank one. The following theorem provides a lower bound for random matrices.

Theorem 1. *If M is a uniformly random $n \times n$ boolean matrix then*

$$\Pr[\mathsf{blocky}(M) \geq \tfrac{n}{4\log(2n)}] \geq 1 - 2^{-\frac{n^2}{2}}.$$

Theorem 1 is proved in Section 2. The lower bound has a factor of $\log n$ compared to the obvious upper bound. This factor turns out to be needed. The blocky rank of *all* boolean matrices is much smaller than n.

Theorem 2. *For every boolean $n \times n$ matrix M,*

$$\mathsf{blocky}(M) \leq O(\tfrac{n \log\log n}{\log n}).$$

The proof of Theorem 2 is algorithmic; see Section 3. We describe an algorithm that gets as input a boolean matrix M and outputs a *decomposition* of M into a sum of $O(\frac{n \log\log n}{\log n})$ blocky matrices.

The two theorems are reminiscent of Shannon's lower bounds and Lupanov's upper bound in the context of boolean circuit complexity. Shannon proved that the circuit complexity of a random n-variate boolean function is at least $\Omega(\frac{2^n}{n})$, while the obvious upper bound is larger. Lupanov proved that in fact the lower bound is sharp; every n-variate boolean function has a boolean circuit of size at most $O(\frac{2^n}{n})$.

The theorems above have the following additional combinatorial interpretation. The *clique cover* number of a graph is the least number of (induced) cliques that are required to cover it. The *intersection* number of a graph is the least k so that the graph can be represented as the intersection graphs over a universe of size k. An *intersection graph* consists of a set $S_v \subseteq [k]$ for each vertex v, so that every two vertices $u \neq v$ are connected by an edge iff $S_v \cap S_u \neq \emptyset$.

Erdös, Goodman and Pósa showed that the clique number is equal to the intersection number [4]. Bollobás, Erdős, Spencer and West proved that the clique cover number of a uniformly random graph on n vertices is at least $\Omega(\frac{n^2}{\log^2 n})$ and at most $O(\frac{n^2 \log\log n}{\log^2 n})$ [3]. Part of their motivation was to understand the interval number of random graphs. Frieze and Reed improved the upper bound to a sharp $O(\frac{n^2}{\log^2 n})$ [5]. Roughly speaking, the connection between their $\frac{n^2}{\log^2 n}$ and our $\frac{n}{\log n}$ is that a typical clique is of size $\log n$, and $\frac{n}{\log n}$ cliques can (sometimes) be glued to a single blocky matrix, so the total number of blocky matrices becomes $\approx \frac{n^2}{\log^2 n} / \frac{n}{\log n}$. It is worth noting that the upper bounds from [3, 5] hold for random graphs and are false for some graphs, whereas our upper bound holds for all matrices.

Let us now make the connection more formal. We work with bipartite graphs, because they correspond to (general) boolean matrices. A *blocky* graph is a bipartite graph that consists of a disjoint union of full bipartite graphs. Equivalently, the adjacency matrix of a blocky graph has blocky rank one. The *blocky cover number* of a bipartite graph is the minimum number of (induced) blocky graphs that are required to cover it.

The blocky cover number can be thought of as a variant of the intersection number. A different way to view an intersection graph is as a map σ that assigns to each vertex v a vector $s(v) \in \{0,1\}^k$, and two vertices $v \neq u$ are connected by an edge if there is $i \in [k]$ so that $s(v)_i = s(u)_i > 0$. We can extend this definition to larger alphabets. An *agreement graph* consists a map σ that assigns to each vertex v a vector $s(v) \in \{0,1,\ldots,L\}^k$, so that every two vertices $v \neq u$ are connected by an edge if there is $i \in [k]$ so that $s(v)_i = s(u)_i > 0$. The integer k is called the *universe size* of the graph (the integer L is not assumed to be bounded).

If G is a bipartite graph, then G has an agreement representation with universe size one iff G is blocky. More generally, the cover number of G using blocky graphs is equal to the least universe size of an agreement representation of G. This is analogous to the fact that the clique cover number is equal to the intersection number [4].

Instead of covering the edges of a graph, we can ask to partition them to structured parts. The *blocky partition number* of a bipartite graph is the minimum number of pairwise edge-disjoint (induced) blocky graphs that are required to cover it. The proof of Theorem 2 actually shows that the blocky partition number of *every* bipartite graph with n vertices is $\tilde{O}(\frac{n}{\log n})$.

1.2. **The greater-than matrix.** A more interesting but often more difficult question is understanding the complexity of specific objects (and not of random objects). We move to investigating the blocky rank of specific matrices.

The first matrix we consider is the $n \times n$ greater-than matrix GT_n defined by $\mathsf{GT}_n(x,y) = 1_{x \leq y}$, where we think of the rows and columns of GT_n as integers in $[n]$. The greater-than matrix is the adjacency matrix of the *half graph*.

Theorem 3. $\mathsf{blocky}(\mathsf{GT}_n) = \Theta(\log n)$.

The upper bound is relatively straightforward and was proved a long time ago in the context of Schur algebras [13]. It actually states that the blocky partition number of the half graph is at most $\lceil \log n \rceil + 1$; see Claim 15 below. In particular, even the *monotone* blocky rank of GT_n is at most order $\log n$ (in monotone ranks, we only allow to use positive coefficients).

A variant of the blocky rank of the greater-than matrix was studied in the context of closure properties of "threshold classes" in machine learning [1, 6]. There are many variants of blocky rank we can study: a monotone version where the linear combination just uses positive numbers, an approximate version where we just need to approximate the target matrix, a signed version where we just need to get the sign pattern correctly, and so forth. Here is a variant that is related to closure properties in machine learning. For a tuple $B = (B_1,\ldots,B_R)$ of $n \times n$ boolean matrices, and a function $F : \{0,1\}^R \to \{0,1\}$, let $F(B)$ be the $n \times n$ matrix obtained by applying F entry-wise: for all i,j,

$$(F(B))_{i,j} = F((B_1)_{i,j},\ldots,(B_R)_{i,j}).$$

Definition. *The* functional *blocky rank* $\mathsf{fun\text{-}blocky}(M)$ *of a boolean matrix M is the minimum number R so that there is a tuple $B = (B_1,\ldots,B_R)$ of blocky matrices and $F : \{0,1\}^R \to \{0,1\}$ so that $M = F(B)$.*

The lower bound $\mathsf{fun\text{-}blocky}(\mathsf{GT}_n) \geq \Omega(\log\log n)$ is implicit in the work of Alon, Beimel, Moran, and Stemmer [1]. The better lower bound $\mathsf{fun\text{-}blocky}(\mathsf{GT}_n) \geq \Omega(\frac{\log n}{\log\log n})$ is implicit in the work of Ghazi, Golowich, Kumar and Manurangsi [6]. These two works consider a more general framework, and their arguments are based on Ramsey theory. And even the stronger lower bound is off by a $\log\log n$ factor. We remove this factor, and get a sharp bound.

Theorem 4. $\mathsf{fun\text{-}blocky}(\mathsf{GT}_n) \geq \frac{1}{8}\log n$.

The lower bound is proved in Section 4. This argument too is related to covering graphs. The Graham–Pollak theorem states that the edges of the full graph on n vertices can not be *partitioned* to less than $n-1$ complete bipartite graphs [7]. In [15], Orlin suggested to study the problem of *covering* the *cocktail party* graph (a full graph minus a perfect matching). Part of his motivation came from computational complexity theory (see Remark 3.7 in his paper). In [8], Gregory and Pullman proved that the clique cover number of the cocktail party graph is $\Theta(\log n)$.

We consider the following bipartite strengthening of their result. The *bipartite cocktail party graph* is the full bipartite graph minus a perfect matching.

Theorem 5. *The blocky cover number of the bipartite cocktail party graph with n vertices on each side is at least $\frac{1}{4}\log n$.*

This is quantitively weaker but more general than the lower bound of Gregory and Pullman. The cocktail party graph contains a copy of the bipartite cocktail party graph. And a clique in the cocktail party graph corresponds to a *connected* blocky graph in that copy. Our lower bound holds also when we are allowed to cover the bipartite cocktail party graph by any blocky graphs (not necessarily connected).

Avishay Tal shared with us the following observation. For every $n \times n$ Boolean matrix M, the functional blocky rank of M is at most $O(\log n)$. The reason is that the matrix B defined by $B_{x,y} = x_i$ for some fixed i is blocky, and with $2\lceil \log n\rceil$ such matrices we can encode both x, y. This shows that, somewhat unusually, GT_n is an explicit matrix of essentially maximum functional blocky rank.

1.3. The inner-product matrix.

The second matrix we consider is the inner-product matrix; let IP_n be the $\{0,1\}^n \times \{0,1\}^n$ matrix defined by $\mathsf{IP}_n(x,y) = \sum_i x_i y_i \,\mathsf{mod}\, 2$. This matrix has been studied in various contexts, like circuit complexity, communication complexity, margin complexity and more. We focus here on its connection to circuit complexity; in particular, to depth-two threshold circuits. There is a long line of research on this topic; see [9, 16, 18, 11, 19, 14, 2] and references within.

A *linear threshold function* (LTF) function is of the form $T(z) = \mathsf{sign}(b + \sum_i a_i z_i)$ for $a_1, \ldots, a_n, b \in \mathbb{Z}$ where sign is 1 on $[0, \infty)$ and 0 on $(-\infty, 0)$. A *majority* gate is a special kind of LTFs in which all constants $a_1, \ldots, a_n$ are in $\{-1, 0, 1\}$. A $\mathsf{MAJ} \circ \mathsf{LTF}$ circuit computes a function of the form $D(z) = m(T_1(z), \ldots, T_s(z))$ where each T_i is an LTF and m is a majority gate. The *size* of the circuit is $|D| = s$.

Hajnal, Maass, Pudlák, Szegedy and Turán proved a lower bound of roughly $2^{n/3}$ for the size of $\mathsf{MAJ} \circ \mathsf{LTF}$ circuits computing the inner product function [9]. Amano constructed a $\mathsf{MAJ} \circ \mathsf{LTF}$ circuit of size $(1.899)^n$ computing inner product [2]. The blocky rank perspective allows to improve the lower bound.

Theorem 6. *If $D \in \mathsf{MAJ} \circ \mathsf{LTF}$ and $D = \mathsf{IP}_n$ then $|D| \geq \Omega(\frac{2^{n/2}}{n})$.*

The theorem is proved in Section 5.1. The proof proceeds by bounding the correlation between IP_n and a threshold gate. An upper bound of $\approx 2^{-n/3}$ on the correlation was proved in [9]. We improve the bound to $\approx 2^{-n/2}$ which is basically sharp; see Claim **??** below.

1.4. Depth-two threshold circuits.

Finally, we describe a general connection between blocky rank and circuit complexity, specifically, depth-two threshold circuits. Proving strong lower bound for $\mathsf{LTF} \circ \mathsf{LTF}$ circuits is a long-standing open problem. Kane and Williams proved the best known lower bound for this model [11]. They proved that the size of every $\mathsf{LTF} \circ \mathsf{LTF}$ circuit computing the n-variate Andreev function must be of size $\Omega(n^{3/2})$.

Roychowdhury, Orlitsky, and Siu observed that we do not even know how to prove lower bounds for $\Sigma \circ \mathsf{LTF}$ circuits, where the upper gate just computes a linear function (with arbitrary coefficients); see [18, 19]. We observe that lower bounds on blocky rank yield circuit lower bounds in this model.

Theorem 7. *Let M be a $\{0,1\}^n \times \{0,1\}^n$ matrix. If $M = \sum_{i=1}^{s} w_i T_i$ where each $w_i \in \mathbb{R}$ and each T_i is an LTF then*

$$s \geq \frac{\mathsf{blocky}(M)}{2(n+1)}.$$

The theorem is proved in Section 5.2. It shows that proving strong lower bound on the blocky rank of explicit boolean matrices might be difficult but rewarding. A similar theorem holds for the *signed* version of blocky rank and general $\mathsf{LTF} \circ \mathsf{LTF}$ circuits.

The theorem suggests that even proving relative weak lower bounds (say, polynomial in n) on the blocky rank of an explicit $2^n \times 2^n$ matrix is interesting. The lower bound from [11] relies on the anti-concentration phenomenon, which does not seem directly relevant to blocky rank. So, even obtaining an $\Omega(n^{5/2})$ lower bound on the blocky rank (which would yield the same circuit lower bound) seems interesting to us.

2. A lower bound for random matrices

The lower bounds follows from a counting argument showing that there are few boolean matrices with low blocky rank.

Lemma 8. *If $V \subset \mathbb{R}^n$ is a linear subspace of dimension k, then*

$$\left|V \cap \{0,1\}^n\right| \leq 2^k.$$

Proof. We can choose a basis for V in echelon form. That is, there are $v_1, \ldots, v_k \in V$ and $i_1 < \ldots < i_k$ in $[n]$ so that $(v_j)_{i_j} = 1$ and $(v_j)_{i_\ell} = 0$ for all $\ell < j$. If $\sum_i a_i v_i \in \{0,1\}^n$, it follows that given $a_1, \ldots, a_i$, there are at most two possible options for a_{i+1}. The total number of possibilities for $a_1, \ldots, a_k$ is at most 2^k. □

Lemma 9. *For $n > 2$, the number of blocky matrices of size $n \times n$ is at most $\frac{1}{2}(2n)^{2n}$.*

Proof. By permuting the rows and columns, every blocky matrix can be brought into a block diagonal form. A matrix in block diagonal form can be represented by two sets $\{i_1 < i_2 < \ldots < i_r\}$ and $\{j_1 < j_2 < \ldots < j_r\}$ in $[n]$ so that the first block is of size $i_1 \cdot j_1$, the second $(i_2 - i_1) \cdot (j_2 - j_1)$ and so on. The case when there are zero rows or columns is encoded by $i_r < n$ or $j_r < n$. There is at most $2^n \cdot 2^n = 2^{2n}$ ways to choose this representation and $n! \cdot n! \leq \frac{1}{2} n^{2n}$ ways to order the rows and columns. $\square$

Proof of Theorem 1. For fixed blocky matrices $B_1, \ldots, B_R$,

$$\left|\mathsf{span}\{B_1, \ldots, B_R\} \cap \{0,1\}^{n \times n}\right| \leq 2^R.$$

The number of boolean blocky matrices of rank at most R is therefore at most $2^R(\frac{1}{2}(2n)^{2n})^R = (2n)^{2nR}$. If $R \leq \frac{n}{4\log(2n)}$ then $(2n)^{2nR} \leq 2^{n^2/2}$. $\square$

3. An upper bound for all matrices

We now provide a non-trivial upper bound for all boolean matrices. We start by dealing with matrices with few ones. For a boolean matrix M, denote by $|M|$ the number of one entries in M.

Lemma 10. *For every $n \times n$ boolean matrix M so that $|M| \leq \frac{n^2}{\log^2 n}$,*

$$\mathsf{blocky}(M) \leq \tfrac{2n}{\log n}.$$

Proof. Denote by $a(M)$ the number of non-zero rows in M, and by $b(M)$ the number of non-zero columns. And let $z(M) = \min\{a(M), b(M)\}$. We can always bound $\mathsf{blocky}(M) \leq z(M)$. So, if $z(M) \leq \frac{2n}{\log n}$, we are done.

We inductively construct a sequence of matrices $M_0 = M, M_1, M_2, \ldots, M_T$ as follows. Assume we have already constructed M_t. If $z(M_T) \leq \frac{n}{\log n}$ then we stop. Otherwise, if $a(M_t) = z(M_t)$, choose a single one-entry in each row in M_t and put all these ones into a blocky matrix B_t. Otherwise, choose a single one-entry in each column in M_t and put all these ones into the blocky matrix B_t. Let $M_{t+1} = M_t - B_t$.

For all $t < T$, we have $|B_t| = z(M_t) > \frac{n}{\log n}$. It follows by induction that for all $t < T$, we have $|M_t| \leq |M_0| - t\frac{n}{\log n}$. Because M_0 has few ones, we know $T \leq \frac{n}{\log n}$. The blocky rank of M_T is at most $z(M_T) \leq \frac{n}{\log n}$. The blocky rank of M is at most $\mathsf{blocky}(M) \leq T + \mathsf{blocky}(M_T)$. $\square$

The upper bound on the blocky rank of general matrices uses the following well-known combinatorial result. A *one-submatrix* of an $n \times n$ boolean M is a submatrix of M so that all of its entries are one. The famous solution of Kövári–Sós–Turán of the Zarankiewicz problem says that if $|M|$ is large then M contains a large one-submatrix [12]. The following lemma is a special case.

Lemma 11. *If n is large enough and M is an $n \times n$ boolean matrix so that $|M| \geq \frac{n^2}{\log^2 n}$, then there exist a one-submatrix of M with at least $\frac{\log n}{5 \log\log n}$ rows and at least $\sqrt{n}$ columns.*

Proof. The reference we use is Lemma 5.4 in [17], which says that if $\epsilon > 0$ and k are so that $\epsilon n \geq 2k$ and $k \leq \frac{\log n}{2\log(2e/\epsilon)}$ then every matrix M with $|M| \geq \epsilon n^2$ contains a one-submatrix with at least k rows and $\sqrt{n}$ columns. Apply this lemma with $\epsilon = \frac{1}{\log^2 n}$. □

Proof of Theorem 2. We describe an algorithm that gets as input an $n \times n$ boolean matrix M, and efficiently decomposes it to blocky matrices. Throughout the proof,

$$k = \left\lceil \frac{\log n}{5\log\log n} \right\rceil \quad \text{and} \quad \ell = \lceil \sqrt{n} \rceil \quad \text{and} \quad h = \lceil \tfrac{2n}{\log n} \rceil \quad \text{and} \quad \tau = \tfrac{n^2}{\log^2 n}.$$

Denote by $b(M)$ the number of non-zero columns in M. Denote by $\Phi(n)$ the maximum blocky rank of an $n \times n$ boolean matrix (we can assume without loss of generality that n is a power of two).

We are going to prove that for all $n' \leq n$,

$$\Phi(n') \leq 2(\tfrac{2n'}{k} + \tfrac{k\ell}{2} + h) + \Phi(\tfrac{n'}{2}). \tag{1}$$

This indeed completes the proof because we can apply the above $J \leq O(\log\log n)$ times for all n' of the form $\frac{n}{2^j}$ between n and h, and deduce that

$$\Phi(n) \leq \tfrac{4n}{k} + \tfrac{4(n/2)}{k} + \ldots + \tfrac{4(n/2^{J-1})}{k} + 2J(\tfrac{k\ell}{2} + h) + \Phi(h) \leq O(\tfrac{n\log\log n}{\log n}).$$

Fix an $n' \times n'$ boolean matrix M. We are going to construct

$$Q \leq \tfrac{2n'}{k} + \tfrac{k\ell}{2} + h$$

blocky matrices $B_1, \ldots, B_Q$ so that the matrix $M^{(Q)} = \sum_{q=1}^{Q} B_q$ is entry-wise at most M (that is, $M^{(Q)}_{ij} \leq M_{ij}$ for all i, j) and the matrix $M - M^{(Q)}$ is so that

$$b(M - M^{(Q)}) \leq \tfrac{n'}{2}.$$

We can apply the same procedure to the matrix $(M - M^{(Q)})^T$ so that we now reach a matrix with at most $\frac{n}{2}$ non-zero rows and columns (where T denotes transposition). This proves (1).

Assume we have already constructed $B_0, \ldots, B_q$ and we want to construct B_{q+1}, where $B_0 = 0$. Let

$$M' = M - \sum_{i \leq q} B_i.$$

Run the following procedure:

(1) $t = 0$ and $N_0 = M'$.
(2) while $|N_t| \geq \tau$ by Lemma 11 define N_{t+1} via

$$N_t = \begin{pmatrix} 1_{k\times\ell} & * \\ * & N_{t+1} \end{pmatrix}.$$

(3) t = t+1 . (end while)

Let T denote the final value of t. Let B be the $n' \times n'$ blocky matrix with the T one-blocks of size $k \times \ell$ that were found above. It could sometimes be the zero matrix. Let L be the $n' \times n'$ matrix defined by

$$L = \begin{pmatrix} 0 & 0 \\ 0 & N_T \end{pmatrix}.$$

There are two options to consider.

The first options is that $T\ell \geq \frac{n'}{2}$: In this case, we set

$$B_{q+1} = B.$$

There are at least $T\ell$ non-zero columns in B and each column has k non-zero entries. So, $|B_{q+1}| \geq \frac{n'k}{2}$. This means that this first option can occur at most $\frac{2n'}{k}$ times because $|M| \leq n'^2$.

The second options is that $T\ell < \frac{n'}{2}$: Denote by B' the matrix obtained from M' by zeroing out every zero row in B. Because the total number of non-zero rows in B is at most $k\frac{n'}{2\ell} \leq \frac{k\ell}{2}$,

$$\mathsf{blocky}(B') \leq \frac{k\ell}{2}.$$

Because $|L| < \tau$, Lemma 10 implies

$$\mathsf{blocky}(L) \leq h.$$

The matrix $M' - L - B'$ satisfies

$$b(M' - L - B') \leq T\ell < \frac{n'}{2}.$$

We thus reduced the number of columns by a factor of two.

□

4. The functional blocky rank of greater-than

Let C be the $n \times n$ matrix with zeros on the diagonal and ones elsewhere; it corresponds to the bipartite cocktail party graph.

Lemma 12. *If $C = \vee(B)$ where $\vee$ denotes the OR function and $B = (B_1, \ldots, B_R)$ is a tuple of $n \times n$ blocky matrices then $R \geq \frac{1}{2}\log n$.*

Proof of Lemma 12. The proof is by induction on n. For $n = 1$, the claim is trivial. For the inductive step, the ones of the matrix B_1 correspond to pairwise disjoint sets $S_1, \ldots, S_A \subseteq [n]$ and $T_1, \ldots, T_A \subseteq [n]$. That is, $(B_1)_{i,j} = 1$ iff $i \in S_a$ and $j \in T_a$ for some $a \in [A]$.

Define two random subsets S and T of $[n]$ as follows. Let $\epsilon_1, \ldots, \epsilon_A$ be i.i.d. uniformly distributed in $\{0, 1\}$. Let S be the complement of $\bigcup_{a:\epsilon_a=1} S_a$ and T be the complement of $\bigcup_{a:\epsilon_a=0} T_a$. Let $I = \{i \in [n] : (i, i) \in S \times T\}$.

The projection of B_1 to $S \times T$ is the zero matrix (with probability one). For each $i \in [n]$, the probability that $(i, i) \in S \times T$ is one quarter, because $(B_1)_{i,i} = 0$. There is a choice for $S \times T$ so that $|I| \geq \frac{n}{4}$.

Let C' be the matrix C after deleting all rows and columns not in I. The matrix C' is a cocktail party matrix of dimension $|I|$, and the matrix B_1 does not contribute to its representation. The inductive hypothesis completes the proof. □

Proof of Theorem 4. Assume that $\mathsf{GT}_n = F(B)$ for $B = (B_1, \ldots, B_R)$ where each B_r is blocky. Assume towards a contradiction that $R < \frac{1}{8}\log n$. There are $m \geq \sqrt{n}$ values of $i \in [n]$ so that the values of $B_{i,i} \in \{0, 1\}^R$ are all equal. Delete the rest $n - m$ rows and columns from GT_n and from B, and focus on the remaining $m \times m$ part. Denote by

G the obtained copy of GT_m, and denote by B' the obtained tuple of matrices so that $G = F(B')$.

There is $f \in \{0,1\}^R$ so that for each $i \in [m]$, we have $B'_{i,i} = f$. Order B' so that the first k entries in f are ones, and the last $R-k$ are zeros.

Claim 13. *For every $i \neq j$ in $[m]$, there is $r > k$ so that (exactly) one of $(B'_r)_{i,j}$ and $(B'_r)_{j,i}$ is one.*

Proof. For each $r \leq k$, because B'_r is blocky and $(B'_r)_{i,i} = (B'_r)_{j,j} = 1$, we know that $(B'_r)_{i,j} = (B'_r)_{j,i}$. Because $G_{i,j} \neq G_{j,i}$, the two lists $((B'_{k+1})_{i,j}, \ldots, (B'_R)_{i,j})$ and $((B'_{k+1})_{j,i}, \ldots, (B'_R)_{j,i})$ must be different. $\square$

The matrix $\vee(B'_{k+1}, \ldots, B'_R, {B'}^T_{k+1}, \ldots, {B'}^T_R)$ is therefore zero on the diagonal and one elsewhere (where T denotes transposition). Lemma 12 implies that $2(R-k) \geq \frac{1}{2}\log m$. $\square$

5. Circuit complexity

5.1. $\mathsf{MAJ} \circ \mathsf{LTF}$ circuits.

In this subsection, we use the blocky rank persepective to prove circuit lower bounds for inner-product.

Definition. *The nuclear norm of the matrix M is*

$$\|M\|_\nu = \inf\Big\{\sum_{i=1}^t p_i : M = \sum_{i=1}^t p_i B_i \quad \forall i \ p_i > 0, \mathsf{rank}(B_i) = 1, \|B_i\|_\infty \leq 1\Big\}.$$

Claim 14. *If B is a blocky matrix then $\|B\|_\nu \leq 1$.*

Proof. The claim follows from the well-known fact that the nuclear norm of the unit matrix is at most one (see e.g. [10]). We include a proof for completeness. It is sufficient to consider $n \times n$ identity matrices for $n+1$ prime. Then, for all $x, y \in [n]$,

$$1_{x=y} = \sum_{z \in \{0,1,\ldots,n\}} \tfrac{1}{n+1} \cdot e^{2\pi i (x-y)z/(n+1)}. \qquad \square$$

Consider the following generalization of LTFs.

Definition. *An $[n] \times [n]$ matrix M is* monotone *if for every x and $y < y'$ in $[n]$,*

$$M_{x,y} \leq M_{x,y'}.$$

Claim 15. *If M is a boolean $n \times n$ monotone matrix then M is a sum of at most $\lceil \log n \rceil + 1$ blocky matrices.*

Proof. Because duplicating rows and columns do not increase the blocky rank, we can consider the matrix GT_n. The blocky rank of GT_n is at most $\lceil \log n \rceil + 1$; see e.g. [13]. We include a proof for completeness. We prove the claim for GT_n for $n = 2^k$. The proof is by induction on k. For the base case $k = 0$, we have $\mathsf{blocky}(G_1) \leq 1$. The matrix $\mathsf{GT}_{2^{k+1}}$ can be written as

$$\mathsf{GT}_{2^{k+1}} = \begin{pmatrix} \mathsf{GT}_{2^k} & J \\ 0 & \mathsf{GT}_{2^k} \end{pmatrix},$$

where J is the all-ones matrix. Let $B_1, \ldots, B_{k+1}$ be the matrices so that $\mathsf{GT}_{2^k} = B_1 + \ldots + B_{k+1}$. We can write $\mathsf{GT}_{2^{k+1}}$ as

$$\mathsf{GT}_{2^{k+1}} = \begin{pmatrix} B_1 & 0 \\ 0 & B_1 \end{pmatrix} + \ldots + \begin{pmatrix} B_{k+1} & 0 \\ 0 & B_{k+1} \end{pmatrix} + \begin{pmatrix} 0 & J \\ 0 & 0 \end{pmatrix}. \qquad \square$$

Corollary 16. *The nuclear norm of an $n \times n$ boolean monotone matrix is at most* $\lceil \log(n) \rceil + 1$

The property of inner-product we rely on is Lindsey's lemma. This was done in many works, including [9]. The proof of the lemma uses the fact that the rows of IP_n are orthogonal using the Cauchy-Schwarz inequality.

Lemma 17. *If M is a $\{0,1\}^n \times \{0,1\}^n$ matrix of rank one so that $\|M\|_\infty \leq 1$ then*

$$\Big| \sum_{x,y \in \{0,1\}^n} (-1)^{\mathsf{IP}_n(x,y)} M(x,y) \Big| \leq 2^{\frac{3n}{2}}.$$

We can conclude the following strengthening of the correlation bound from [9].

Lemma 18. *If T is an LTF then*

$$\Big| \sum_{x,y \in \{0,1\}^n} (-1)^{IP_n(x,y)} T(x,y) \Big| \leq (n+1) 2^{\frac{3n}{2}}.$$

Proof. The matrix T is monotone (up to a permutation of the rows and columns). Corollary 16 bounds the nuclear norm of T from above; we can write

$$T = \sum_i p_i B_i$$

where each $p_i > 0$, where each B_i is of rank one and $\|B_i\|_\infty \leq 1$, and where $\sum_i p_i \leq (n+1)$. Lemma 17 implies

$$\begin{aligned} \Big| \sum_{x,y \in \{0,1\}^n} (-1)^{IP_n(x,y)} T(x,y) \Big| &= \Big| \sum_{x,y \in \{0,1\}^n} (-1)^{IP_n(x,y)} \sum_i p_i B_i(x,y) \Big| \\ &\leq \sum_i p_i \Big| \sum_{x,y \in \{0,1\}^n} (-1)^{IP_n(x,y)} B_i(x,y) \Big| \\ &\leq (n+1) 2^{\frac{3n}{2}}. \qquad \square \end{aligned}$$

Proof of Theorem 6. Assume that $\mathsf{IP}_n(x,y) = \mathsf{sign}(-b + \sum_{i=1}^s w_i T_i(x,y))$ where each T_i is an LTF and $w_i \in \{-1,0,1\}$. It follows that $|b| \leq s$ because otherwise the right hand side is constant. For all x,y,

$$(-1)^{1+\mathsf{IP}_n(x,y)} \Big(-1 + 2\Big(-b + \sum_i w_i T_i(x,y) \Big) \Big) \geq 1.$$

Summing over all x,y,

$$\Big| \sum_{x,y} (-1)^{\mathsf{IP}_n(x,y)} \Big(-1 + 2\Big(-b + \sum_i w_i T_i(x,y) \Big) \Big) \Big| \geq 2^{2n}.$$

Lemma 18 implies that

$$\begin{aligned}&\Big|\sum_{x,y}(-1)^{\mathsf{IP}_n(x,y)}\Big(-1+2\Big(-b+\sum_i w_iT_i(x,y)\Big)\Big)\Big|\\&\leq 2(s+1)2^n+2\cdot\sum_i w_i\Big|\sum_{x,y}(-1)^{\mathsf{IP}_n(x,y)}T_i(x,y)\Big|\\&\leq s\cdot 2(n+2)2^{\frac{3n}{2}}.\end{aligned}$$ □

5.2. $\Sigma\circ\mathsf{LTF}$ **circuits.** In this subsection, we show that blocky rank lower bounds imply circuit lower bounds.

Proof of Theorem 7. If $M=\sum_{i=1}^{s}w_iT_i$ then by Claim 15 we have

$$\mathsf{blocky}(M)\leq\sum_i\mathsf{blocky}(T_i)\leq s\cdot(n+1).$$ □

References

[1] N. Alon, A. Beimel, S. Moran, and U. Stemmer. Closure properties for private classification and online prediction. In *COLT*, pages 119–152, 2020.

[2] K. Amano. On the size of depth-two threshold circuits for the inner product mod 2 function. In *LATA*, pages 235–247, 2020.

[3] B. Bollobás, P. Erdős, J. Spencer, and D. B. West. Clique coverings of the edges of a random graph. *Combinatorica*, 13(1):1–5, 1993.

[4] P. Erdös, A. W. Goodman, and L. Pósa. The representation of a graph by set intersections. *Canadian Journal of Mathematics*, 18:106–112, 1966.

[5] A. Frieze and B. Reed. Covering the edges of a random graph by cliques. *Combinatorica*, 15(4):489–497, 1995.

[6] B. Ghazi, N. Golowich, R. Kumar, and P. Manurangsi. Near-tight closure bounds for the littlestone and threshold dimensions. In *ALT*, pages 686–696, 2021.

[7] R. L. Graham and H. O. Pollak. On the addressing problem for loop switching. *The Bell system technical journal*, 50(8):2495–2519, 1971.

[8] D. A. Gregory and N. J. Pullman. On a clique covering problem of orlin. *Discrete Mathematics*, 41(1):97–99, 1982.

[9] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46(2):129–154, 1993.

[10] L. Hambardzumyan, H. Hatami, and P. Hatami. Dimension-free bounds and structural results in communication complexity. *ECCC*, TR21-066, 2021.

[11] D. M. Kane and R. Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *STOC*, pages 633–643, 2016.

[12] P. Kővári, V. T Sós, and P. Turán. On a problem of zarankiewicz. In *Colloquium Mathematicum*, volume 3, pages 50–57. Polska Akademia Nauk, 1954.

[13] S. Kwapień and A. Pełczyński. The main triangle projection in matrix spaces and its applications. *Studia Mathematica*, 34(1):43–67, 1970.

[14] M. Mahajan. Depth-2 threshold circuits. *Resonance*, 24(3):371–380, 2019.

[15] J. Orlin. Contentment in graph theory: covering graphs with cliques. In *Indagationes Mathematicae*, volume 80, pages 406–424, 1977.

[16] R. Paturi and M. E. Saks. Approximating threshold circuits by rational functions. *Information and Computation*, 112(2):257–272, 1994.

[17] A. Rao and A. Yehudayoff. *Communication complexity and applications*. Cambridge University Press, 2020.

[18] V. Roychowdhury, K.-Y. Siu, and A. Orlitsky. Neural models and spectral methods. In *Theoretical Advances in Neural Computation and Learning*, pages 3–36. Springer, 1994.

[19] R. R. Williams. Limits on representing boolean functions by linear combinations of simple functions: thresholds, relus, and low-degree polynomials. In *CCC*, pages 1–24, 2018.

Department of Mathematics, Technion-IIT
Email address: kspy-d@campus.technion.ac.il

Department of Mathematics, Technion-IIT
Email address: amir.yehudayoff@gmail.com

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il