ISRAEL JOURNAL OF MATHEMATICS **TBD** (XXXX), 1–37
DOI: 10.1007/s000000000000000000000000

AROUND THE LOG-RANK CONJECTURE

BY

TROY LEE

Centre for Quantum Software and Information, University of Technology Sydney
P.O.B. 123, Ultimo, NSW 2007, Australia
e-mail: troyjlee@gmail.com

AND

ADI SHRAIBMAN

School of Computer Science, The Academic College of Tel-Aviv-Yaffo
P.O.B. 8401, Tel-Aviv-Yaffo 6818211, Israel
e-mail: adish@mta.ac.il

ABSTRACT

The log-rank conjecture states that the communication complexity of a boolean matrix A is bounded by a polynomial in the log of the rank of A. Equivalently, it says that the chromatic number of a graph is bounded quasi-polynomially in the rank of its adjacency matrix. This old conjecture is well known among computer scientists and mathematicians, but despite extensive work it is still wide open. We survey results relating to the log-rank conjecture, describing the current state of affairs and collecting related questions. Most of the results we discuss are well known, but some points of view are new. One of our hopes is to paint a path to the log-rank conjecture that is made of a series of smaller questions, which might be more feasible to tackle.

Received June 1, 2022 and in revised form January 5, 2023

1. Introduction

The notorious log-rank conjecture asserts that communication complexity and the log of the rank of a boolean matrix (that is, a matrix whose entries are either 0 or 1) are polynomially related. Let A be a boolean matrix, and denote by $r(A)$ the rank of A over the reals. Let the **binary rank** of A, denoted $p(A)$, be the minimum number p such that there exist boolean matrices $\{B_i\}_{i=1}^p$ satisfying $r(B_i) = 1$ and $\sum_i B_i = A$. It follows from sub-additivity of rank that $r(A) \leq p(A)$. The log-rank conjecture says that there exists a constant c such that

$$p(A) \leq 2^{\log^c r(A)}.$$

Observe that the binary rank of A is equal to the minimal size of a partition of the 1-entries of A into monochromatic sub-matrices (that is, rank-1 sub-matrices). For this reason binary rank is sometimes also called the **partition number**. The binary rank/partition number of a boolean matrix A and the communication complexity of A, denoted $cc(A)$, are closely related,

$$p(A) \leq 2^{cc(A)} \leq 2^{\log^2 p(A)}. \tag{1}$$

Therefore, to keep things simple, we do not define communication complexity here and instead make do with the definition of binary rank. The interested reader is referred to [25] to learn the basics of communication complexity and its relation to the partition number. The inequalities in Equation (1) are both tight by the way, see [17].

The definition of partition number is easily extended to non-negative integer matrices: it is the minimum number p such that there exist non-negative integer matrices $\{B_i\}_{i=1}^p$ satisfying $r(B_i) = 1$ and $\sum_i B_i = A$. When we do not restrict to boolean matrices though, it is easy to come up with a matrix A exhibiting a large gap between $r(A)$ and $p(A)$. Let $A = (a_{i,j})$ be the $n \times n$ matrix satisfying $a_{i,j} = i + j$, where indexing begins from 1. It is not hard to verify that the rank of this **Hankel matrix** is equal to 2. The partition number on the other hand is n. All entries on the anti-diagonal of A are equal to $n + 1$. Any 2×2 submatrix of A including two anti-diagonal entries is necessarily of rank 2 as it cannot have zero determinant. Thus all anti-diagonal elements must be in different sets of the partition. Hence, the restriction to boolean matrices is crucial to the log-rank conjecture. We expand on this issue and also discuss other variations of the log-rank conjecture in Section 5.

In the communication complexity context, the log-rank conjecture was first phrased (as a question) by Lovász and Saks [33, 34]. The reason they asked this question is that for the family of matrices they considered, the log-rank conjecture held. But, as Lovász and Saks observed, the log-rank conjecture is equivalent to a conjecture in graph theory suggested (with somewhat different parameters) by van Nuffelen [38] and Fajtlowicz [14]. The latter work is based on a computer program, called Graffiti, that makes graph-theoretic conjectures. Both van Nuffelen and Graffiti (Fajtlowicz) conjectured that the chromatic number of a graph is bounded from above by the rank of its adjacency matrix. We describe the connection between the log-rank conjecture and the graph theoretic conjecture in Section 2, where we also discuss other aspects of the log-rank conjecture that are specific to graphs.

As mentioned above, the conjecture of van Nuffelen and Fajtlowicz was that $\chi(G) \leq r(A_G)$, where $\chi(G)$ is the chromatic number of the graph G and A_G is its adjacency matrix. This conjecture was too strong, and a graph G with $\chi(G)=32$ and $r(A_G)=29$ was found by Alon and Seymour [7]. Razborov [40] proved that the gap between the chromatic number of a graph and the rank of its adjacency matrix can be superlinear. An improved separation was then given by Raz and Spieker [39]. They gave an infinite family of matrices for which $cc(A) \geq \log r(A) \log\log\log r(A)$. Nisan and Wigderson [37] constructed a boolean matrix A for which $cc(A) = \Omega(\log^{\log_2 3} r(A))$ implying that the constant c in the log-rank conjecture must be at least $\log_2 3 \approx 1.58$. Kushilevitz, in an unpublished work, gave an improved base-case example using the same construction to show that $c \geq \log_3 6$ (see [37] for details). The best known separation is given by Göös, Pitassi and Watson [17], who showed that the constant in the log-rank conjecture must be at least 2. We give the details of the best known gaps in Section 4.

The known upper bounds in the log-rank conjecture are very far from the best known lower bounds. It is not hard to see that $p(A) \leq 2^{r(A)}$ for every boolean matrix A, and that $\chi(G) \leq 2^{r(A_G)}$ for every graph G. Kotlov and Lovász [23] proved that the number of distinct rows/columns in the adjacency matrix of a graph G is bounded by $O(2^{r(A_G)/2})$. It follows that $\chi(G) \leq O(2^{r(A_G)/2})$ as the chromatic number is bounded by the number of distinct rows in A_G. Kotlov and Lovász showed that their bound is best possible, exhibiting a graph G with no **twin vertices**, that is vertices corresponding to identical rows in the adjacency matrix, containing $\Omega(2^{r(A_G)/2})$ vertices. Kotlov [22] therefore used a different method to improve the upper bound to $\chi(G) \leq O(r(\frac{4}{3})^r)$, where $r = r(A_G)$. The next result is due to Lovett [35] who proved a significantly improved bound $p(A) \leq 2^{O(\sqrt{r(A)} \log r(A))}$, for every boolean matrix A. Lovett's bound is inspired by previous works with Ben-Sasson and Ron-Zewi [11] and with Gavinsky [16]. Its starting point is a beautiful upper bound of Nisan and Wigderson [37] depending on the largest size of a rank-1 submatrix, and on rank. These upper bounds are the subject of Section 3.

To sum up, the goal of this survey is to collect results about the log-rank conjecture in order to highlight possible research directions. We describe known upper bounds in Section 3, and known lower bounds in Section 4. In Section 2 we focus on aspects of the log-rank conjecture that relate to graphs. Here we describe some questions that were not considered before. In Section 5 we consider variants of the log-rank conjecture. In this section we try to identify where the log-rank conjecture breaks when going to integer matrices, and attempt to stretch a line between what is known and what is still not known.

2. Rank and chromatic number

As mentioned, the graph theoretic question, relating chromatic number and rank, preceded the equivalent log-rank conjecture in communication complexity. But for some reason this point of view received less attention. The communication complexity literature contains several questions that are derived from the log-rank conjecture, e.g., the log-rank conjecture for special families of matrices like the ones related to composed functions. In Section 2.2 we suggest some graph theoretic questions related to the log-rank conjecture, that seem natural and interesting in their own right. The simplest open question is: what is the minimal multiplicity of -1 as an eigenvalue of the adjacency matrix of a triangle-free graph? This question, and generalizations of it, are related to the log-rank conjecture via the fooling set method.

This section is somewhat independent from the rest of this survey, but it contains some basic results that are used later, which is why it appears first. We start in Section 2.1 by giving the proof of the equivalence to the log-rank conjecture in communication complexity, as it appears in [34], adding small details when necessary.

2.1. The equivalence to the log-rank conjecture. The equivalence of the two conjectures uses the cover number of a matrix. Given a boolean matrix A, the **cover number** of A, denoted by $\chi(A)$, is the minimum number χ such that there exist boolean matrices $\{B_i\}_{i=1}^{\chi}$ with $r(B_i) = 1$ and $B_i \leq A$ for each i, satisfying $\sum_i B_i \geq A$. Here the inequalities are entry-wise.

The cover number is of course upper bounded by the partition number, and usually the gap can be very large. But in this case either the cover number of A, or the cover number of its complement $\bar{A}$ ($\bar{A} = J - A$ where J is the all-ones matrix), cannot differ much from the partition number.

Lemma 2.1 (Halstenberg and Reischuk [18, Theorem 1]): *For every boolean matrix A it holds that*

$$\log p(A) = \log \chi(A) \log \chi(\bar{A})(1 + o(1)).$$

A version of the above bound, with slightly weaker estimates, was first proved by Aho, Ullman and Yannakakis [3].

Back to the equivalence of the log-rank conjecture and its graph theoretic counterpart. Denote by A_G the adjacency matrix of a graph G.

Lemma 2.2 (Lovász and Saks [34, Question 8.2]): *The following statements are equivalent:*

(1) *There is a constant c such that for every graph G*

$$\chi(G) \leq 2^{\log^c r(A_G)}.$$

(2) *There is a constant c' such that for every boolean matrix A*

$$\chi(A) \leq 2^{\log^{c'} r(A)}.$$

(3) *The log-rank conjecture.*

Proof. We start with (2) implies (1). Let G be an n-vertex simple graph, and denote by $A = A_G$ the adjacency matrix of G. Then clearly the chromatic number of G satisfies

$$\chi(G) \leq \chi(\bar{A}) \leq 2^{\log^c r(\bar{A})} \leq 2^{\log^c (r(A)+1)}.$$

For the other direction, (1) implies (2), let A be a boolean matrix. Define a graph $G = G_A = (V_A, E_A)$, where V_A is the subset of row, column index pairs (i, j) for which $a_{i,j} = 1$. Two vertices $v_1 = (i_1, j_1)$ and $v_2 = (i_2, j_2)$ are adjacent if $a_{i_1,j_2} = 1$ and also $a_{i_2,j_1} = 1$. That is, the vertices are adjacent if the corresponding two entries "span" a rank-1 submatrix of A. Observe that G has a self loop on each of its vertices, and that the maximal (w.r.t. inclusion) cliques of G_A are exactly the maximal rank-1 sub-matrices of A.

Therefore, the cover number of A satisfies

$$\chi(A) = \chi(\bar{G}) \leq 2^{\log^c r(A_{\bar{G}})} \leq 2^{\log^c (r(A_G)+1)}.$$

Here $\bar{G}$ is the complement of the graph G. Since G has self loops, we have that the adjacency matrix of the complement graph is the complement of the adjacency matrix A_G, that is $A_{\bar{G}} = \bar{A}_G$. As a consequence, the rank of A_G and the rank of $A_{\bar{G}}$ differ by at most 1.

But, A_G is a submatrix of the tensor product $A \otimes A^t$ (simply take rows/columns that correspond to pairs (i,j) for which $a_{i,j} = 1$) and therefore

$$\chi(A) \leq 2^{\log^c (r^2(A)+1)}.$$

It is left to show that (2) is equivalent to (3). Assume the log-rank conjecture is true, then for every boolean matrix A there is a constant c such that

$$\chi(A) \leq p(A) \leq 2^{\log^c r(A)}.$$

For the other direction, assume (2) holds. Then there is a constant c such that for every boolean matrix A, $\chi(A) \leq 2^{\log^c r(A)}$. Applying the same reasoning to the complement $\bar{A}$ we get that also $\chi(\bar{A}) \leq 2^{\log^c r(\bar{A})} \leq 2^{\log^c (r(A)+1)}$. By Lemma 2.1 we conclude that

$$\log p(A) \leq O(\log \chi(A) \log \chi(\bar{A})) \leq O((2 \log r(A))^{2c}). \qquad \blacksquare$$

2.2. The extended fooling set method. Let A be an $m \times n$ boolean matrix. A **fooling set** for A is a set of row, column index pairs $\{(i_1, j_1), \dots, (i_k, j_k)\}$ such that

(1) $a_{i_t, j_t} = 1$ for all $t = 1, \dots, k$,
(2) $a_{i_s, j_t} \cdot a_{i_t, j_s} = 0$ for all $s \neq t$.

Note that these conditions in particular imply that $i_s \neq i_t$ and $j_s \neq j_t$ for all $s \neq t$.

If A has a fooling set of size k, then there is a $k \times k$ submatrix B of A that (after a permutation of rows and columns) has ones on the main diagonal and satisfies $B \circ B^t = \mathbf{I}_k$, where $\circ$ is the Hadamard (entry-wise) product, and $\mathbf{I}_k$ is the $k \times k$ identity matrix. As $k = r(B \circ B^t) \leq r(B)^2$, if A has a fooling set of size k then $r(A) \geq \sqrt{k}$. Thus the rank bound is never much worse than the fooling set bound. Amazingly, this simple argument is nearly tight. Shigeta and Amano [44] give a construction of $n \times n$ boolean matrices with a fooling set of size n and rank $n^{1/2+o(1)}$.

Let $\mathrm{fs}(A)$ be the maximum k such that A has a fooling set of size k. It is not hard to verify that $\mathrm{fs}(A) \le \chi(A) \le p(A)$: in terms of G_A, a fooling set provides an independent set in G_A of cardinality k, therefore, the chromatic number of the complement of G_A (which is equal to $\chi(A)$) is at least k. The fooling-set method is one of the first lower bound techniques for communication complexity. This method does not always give good lower bounds [13]. In fact, for a random $n \times n$ boolean matrix A it holds that $\mathrm{fs}(A) = O(\log n)$ with high probability, whereas $r(A) = \chi(A) = n$ almost surely. There are also explicit examples with an exponential gap. For example, let $\mathbb{F}_2^t = \{v_1, \ldots, v_{2^t}\}$ be the vectors in the t-dimensional vector space over $\mathbb{F}_2$, and consider the $2^t \times 2^t$ matrix H whose (i,j)-th entry is $v_i v_j^T$. That is, the entries of H are the inner products of vectors in $\mathbb{F}_2^t$. It is not hard to verify that $r(H) \ge 2^t - 1$. In fact, if we change each 0-entry in H to 1 and each 1-entry to -1 then the resulting matrix is a **Hadamard matrix**, namely a square matrix whose entries are ± 1 and whose rows are mutually orthogonal. On the other hand $\mathrm{fs}(H) \le t^2$, since the rank over $\mathbb{F}_2$ of H is at most t, and the size of any fooling set is bounded from above by the square of the rank over any field.

Dietzfelbinger et al. suggested a strengthening of the fooling-set method called the extended fooling set method [13]. Let $G = G_A$. As mentioned above, a fooling set provides a clique in $\bar{G}$ of size k, and therefore a lower bound on the chromatic number of $\bar{G}$. This method can be extended by looking for a subgraph of $\bar{G}$ with k vertices whose independence number is upper bounded by some integer d, providing the lower bound k/d for the chromatic number. A **fooling set of order** d for a boolean matrix A is therefore a subset of the 1-entries of A satisfying that no $d+1$ of them belong to a rank-1 submatrix of A, together.

Formally, a fooling set of order d for an $m \times n$ boolean matrix A is a set of row, column index pairs $F = \{(i_1, j_1), \ldots, (i_k, j_k)\}$ such that

(1) $a_{i_t, j_t} = 1$ for all $t = 1, \ldots, k$,

(2) for any selection of $d+1$ index pairs from F, the submatrix of A that is spanned by them has rank larger than 1.

We define $\mathrm{fs}_d(A)$ as the maximum cardinality of a fooling set of order d for A. Note that $\mathrm{fs}(A) = \mathrm{fs}_1(A)$, and that $\mathrm{fs}_d(A)/d \le \chi(A)$ for every $d \ge 1$.

If for some d we find a boolean matrix A with a large gap between rank and $\mathrm{fs}_d(A)/d$, then this also exhibits a gap between rank and $\chi(A)$, and hence also $p(A)$. This is therefore an approach that can possibly refute the log-rank conjecture. Interestingly, as shown by the following theorem, if no such refutation can be found, the log-rank conjecture does hold, at least up to log factors.

Theorem 2.3 (Dietzfelbinger et al. [13, Theorem 1.9]): *For every $m\times n$ boolean matrix A there is natural number d such that*

$$\chi(A) = O\Big(\frac{\mathrm{fs}_d(A)}{d}\log(mn)\Big).$$

The above theorem, combined with Lemma 2.2, implies that the log-rank conjecture is (almost) equivalent to the statement: there is a constant c such that for every boolean matrix A, and every d, it holds that $\frac{\mathrm{fs}_d(A)}{d} \le 2^{\log^c r(A)}$.

It is therefore natural to ask how the extended fooling-set method relates to rank. A nice aspect of this question is that it can be formulated in the language of forbidden minors in graphs. In addition, we can break it down by considering a fixed d at a time. The question is interesting already for $d = 2$.

Let $g(d,k)$ be the minimum rank of a matrix spanned by a fooling-set of order d and cardinality k. The relationship between fs_d and rank is determined by the behaviour of $g(d,k)$. Since we are interested in asymptotic behaviour, we can replace $g(d,k)$ by a symmetric version, involving only graphs. Denote by $f(d,k)$ the minimum rank of the adjacency matrix of a $(d+1)$-clique free graph on k vertices, with self loops. Obviously $g(d,k) \le f(d,k)$. Since $B \circ B^t$ is symmetric and $r(B \circ B^t) \le r(B)^2$, we also have that $f(d,k) \le g(d,k)^2$. Note that if B is a matrix spanned by a fooling-set of order d and cardinality k, then $B \circ B^t$ is the adjacency matrix of a $(d+1)$-clique-free graph on k vertices, with self loops.

The goal is therefore to study $f(d,k)$. Put as a question, it is

Question 2.4: Fix a natural number $d \ge 1$, let G be a k-vertex graph with self loops, and let A_G be the adjacency matrix of G. What is the minimum possible rank of A_G, if G does not contain a clique of size $d+1$?

For $d = 1$ (i.e., the fooling-set method) the question is trivial. But for $d = 2$ it is already interesting and resolving it might hint on whether the log-rank conjecture is true or false. The question becomes:

Question 2.5: What is the minimum rank of the adjacency matrix of a triangle-free graph on k vertices, with self loops?

Let A_G be the adjacency matrix of a triangle-free simple graph on k vertices, then $\mathbf{I}_k + A_G$ is the adjacency matrix of a triangle-free k-vertex graph with self loops. Note that λ is an eigenvalue of $\mathbf{I}_k + A_G$ if and only if $\lambda - 1$ is an eigenvalue of A_G, simply because every vector is an eigenvector of $\mathbf{I}_k$ for the eigenvalue 1. In particular, the multiplicity of 0 as an eigenvalue of $\mathbf{I}_k + A_G$ is equal to the multiplicity of -1 as an eigenvalue of A_G. Therefore, as the rank is equal to the number of non-zero eigenvalues, the matrix $\mathbf{I}_k + A_G$ has rank r if and only if -1 is an eigenvalue of A_G with multiplicity $k - r$. Therefore, the question is alternatively:

Question 2.6: What is the maximum multiplicity in which the value -1 can repeat as the eigenvalue of the adjacency matrix of a triangle-free simple graph on k vertices?

Note that if G has an independent set of size α then $r(I_n + A_G) \geq \alpha$ because then $I_n + A_G$ contains an identity matrix of size α. Turán showed that an n-vertex graph with average degree δ has an independent set of size at least $n/(\delta + 1)$ [47]. A delightful one-page proof of this fact is given in the book The Probabilistic Method [8], after Chapter 6. Turán's theorem can be used to give the following simple bound on $f(2, k)$.

LEMMA 2.7: *For $k > 2$ it holds that $\sqrt{k} \leq f(2, k) \leq \lceil k/2 \rceil$.*

Proof. For the upper bound take a matching of size $\lfloor k/2 \rfloor$, plus maybe an isolated vertex if k is odd.

For the lower bound, let A be the adjacency matrix of a triangle-free graph $G = (V, E)$ on k vertices. By Turán's theorem, $r(I_k + A) \geq n/(\delta + 1)$, where δ is the average degree. Let Δ be the maximum degree of G. We will assume that $\Delta > 1$ as otherwise G is a matching (plus, perhaps, isolated vertices) and we have the lower bound of $\lceil k/2 \rceil \geq \sqrt{k}$ for $k > 2$.

As G is triangle free, the neighbors of any vertex form an independent set. This means that G contains $K_{1,\Delta}$, the complete bipartite graph with color classes of size 1 and Δ, as an induced subgraph. It is easily verified that

$$r(I_{\Delta+1} + A_{K_{1,\Delta}}) = \Delta + 1$$

for $\Delta > 1$. Thus we have

$$r(I_k + A) \geq \max\left\{\frac{k}{\delta + 1}, \Delta + 1\right\} \geq \sqrt{k}. \qquad \blacksquare$$

A slight improvement to the lower bound on $f(2,k)$ in Lemma 2.7 follows from a connection with the Ramsey number $R(3,t)$, the minimum number n such that any n vertex graph contains either a triangle or an independent set of size t. Ajtai, Komlós and Szemerédi [4] show that a k vertex triangle-free graph with average degree δ has an independent set of size at least $0.01(k/\delta)\ln(\delta)$, which implies

$$f(2,k) = \Omega(\sqrt{k\ln(k)}).$$

Ramsey bounds can also be used to generalize Lemma 2.7 to any d. To get a prettier bound, we first show the following little lemma.

Lemma 2.8: *Let G be an n-vertex graph with no $(d+1)$-clique. If G has an independent set of size α then*

$$r(I_n + A_G) \geq \min\{\lfloor n/d \rfloor, \alpha + 1\}.$$

Proof. Let S be a maximal independent set in G of size at least α. If $|S| \geq \lfloor n/d \rfloor$, then we are done, so suppose $|S| < \lfloor n/d \rfloor$.

Every vertex in $\bar{S}$ has a neighbor in S, as S is maximal. Since $|S| < \lfloor n/d \rfloor$ we have $|\bar{S}| > (d-1)|S|$ and there is a vertex $v \in S$ which has at least d neighbors in $\bar{S}$.

Let T be the set of these neighbors. As G does not have a $(d+1)$-clique and every vertex in T is adjacent to v, there must be $u_1, u_2 \in T$ that are not adjacent.

Thus G contains an induced bipartite graph where one color class is S and the other is $\{u_1, u_2\}$. Let

$$B = \begin{bmatrix} \mathbf{0}_{|S|\times|S|} & C^t \\ C & \mathbf{0}_{2\times 2} \end{bmatrix}$$

be the adjacency matrix of this bipartite graph. The result will follow if we show that the multiplicity of -1 as an eigenvalue of B is at most 1, as then

$$r(I_{|S|+2} + B) \geq |S| + 1.$$

The multiplicity of -1 as an eigenvalue of B is equal to the multiplicity of 1 as a singular value of C, which in turn is equal to the multiplicity of 1 as an eigenvalue of $D = CC^t$.

Letting C_1, C_2 denote the rows of C we have

$$D = \begin{bmatrix} \|C_1\|^2 & \langle C_1, C_2\rangle \\ \langle C_2, C_1\rangle & \|C_2\|^2 \end{bmatrix}.$$

Assume for contradiction that both eigenvalues of D are 1. Then the trace of D is 2 which means that $\|C_1\|^2 = \|C_2\|^2 = 1$, as this is the only way for the sum of the squares of two integers to be 2. We further know that $\langle C_1, C_2\rangle \neq 0$, as u_1, u_2 are both adjacent to v. Subject to $\|C_1\|^2 = \|C_2\|^2 = 1$ we must then have $\langle C_1, C_2\rangle = 1$, implying that the determinant of D is 0, a contradiction. ∎

Lemma 2.9: *For $k > d$ it holds that*

$$\Omega(k^{1/d}(\ln(k)/d)^{(d-1)/d}) = f(d,k) \leq \lceil k/d \rceil.$$

Proof. For the upper bound take the union of d-cliques.

For the lower bound, let G be a k-vertex graph without a $(d+1)$-clique such that $r(I_k + A_G) = f(d,k)$. Let α be the size of a largest independent set in G. By Lemma 2.8 we know that $f(d,k) \geq \min\{\lfloor k/d \rfloor, \alpha + 1\}$. If the minimum is achived by $\lfloor k/d \rfloor$ then we are done, so suppose we just know $f(d,k) \geq \alpha + 1$. This means that G contains neither a $(d+1)$-clique nor an independent set of size $f(d,k)$, which means $k < R(d+1, f(d,k))$. Ajtai et al. [4] show that $R(d+1,t) \leq 5000^{d+1} t^d / \ln(t)^{d-1}$. Thus

$$f(d,k) = \Omega(k^{1/d}(\ln(k)/d)^{(d-1)/d}). \qquad ∎$$

Observe that the log-rank conjecture would imply that $f(d,k)$ is closer to the upper bound than the lower bound in the statement of Lemma 2.9, as d grows. If the log-rank conjecture is true then there is a constant c such that

$$f(d,k) \geq 2^{\sqrt[c]{\log k/d}},$$

for every $d \geq 1$. And if the above bound holds then the log-rank conjecture holds up to a $\log\log$ additive factor. The log-rank conjecture therefore comes down to the existence (or non-existence) of a graph without $(d+1)$-cliques, for some d, and rank much lower than $2^{\sqrt[c]{\log k/d}}$. Given that the best upper bound we currently have is k/d, achieved by taking the union of d-cliques, this is one point of view where the log-rank conjecture actually seems plausible, unlike many other perspectives where it usually feels the other way around.

3. Upper bounds

Although the gap in our knowledge regarding the log-rank conjecture is very wide, there are interesting upper bounds. Here is a partial list of older results, one of which we already mentioned:

(1) $cc(A) \leq \log \chi(A) \log \chi(\bar{A})(1 + o(1))$ (Theorem 1 [18]),
(2) $cc(A) \leq (\log \chi(A) + 2)(\log r(A) + 1)$ (Theorem 2.8 [34]),
(3) $cc(A) \leq \mathrm{fs}(A)(\chi(\bar{A}) + 1)$ (Theorem 3.6 [32]).

Gavinsky and Lovett [16] augmented the above repertoire of upper bounds on deterministic communication complexity. They proved that when the rank of the matrix is low, deterministic communication complexity of a matrix A is bounded by randomized communication complexity, and also other relaxed complexity measures.

All the upper bounds mentioned above, the older results and the newer ones, can be derived from an upper bound by Nisan and Wigderson [37], described in Section 3.1. This upper bound is probably what led Nisan and Wigderson to state the log-rank conjecture, which previously only appeared as a question by Lovász and Saks [33, 34]. It is also the starting point for Lovett's improved bound in terms of rank [35], described in Section 3.2. It is used to prove other improved bounds, in terms of relaxed notions of rank, described in Section 3.3.

3.1. The basic upper bound. Known upper bounds on communication complexity, involve two complexity measures, one serves as a potential function (e.g., $\mathrm{fs}(A)$ or $r(A)$), and the other ($\chi(A)$ or $\chi(\bar{A})$) serves as a pool of large submatrices of small rank (either 1 or 0). The log-rank conjecture can be seen as asking whether rank alone can play both roles. In the protocol of Nisan and Wigderson, rank serves like before as a kind of a potential function, while $\chi(A)$ and $\chi(\bar{A})$ on the other hand are replaced by a more rudimentary measure, involving only the size of a largest submatrix of rank at most 1.

For an integer matrix A, denote by $\alpha(A)$ the minimum integer d such that every $s \times t$ submatrix of A has a submatrix of rank at most 1 and size at least $2^{-d}st$.

Theorem 3.1 (Nisan and Wigderson [37, Theorem 2]): *Let A be an $m \times n$ integer matrix. If the entries of A contain at most M different integer values then*

$$\log p(A) = O(\log r(A)(\alpha(A) + \log r(A) + \log \log M)).$$

Proof. Let B be a submatrix of A of rank at most 1 and size at least $2^{-\alpha(A)}mn$. Divide A into four disjoint sub-matrices B, C, D, E, where C is the submatrix of A sharing the rows of B, and D is the submatrix with the same columns set as B.

It holds that $r(C) + r(D) \leq r(A) + 1$. We construct our partition recursively. If $r(C) \leq r(A)/2 + 1$ we start with the partition containing two submatrices, the one containing C and B and the one with D and E. Otherwise, it must be that $r(D) \leq r(A)/2 + 1$ and therefore we choose instead the submatrix containing D and B and the one with C and E. In both cases, we start with two submatrices, one of which has rank at most $r(A)/2 + 2$ and the other has at most $(1 - 2^{-\alpha(A)})mn$ entries. We then recursively refine the submatrices in the partition, in a similar way, until all the submatrices are of rank at most 1. The resulting recursion tree has the following properties:

(1) The recursion tree is a binary tree whose nodes are labelled by a pair of labels $(\text{rank}, \text{size})$.
(2) For every inner node v, the rank label decreases by roughly a factor of 2 when moving to the left child of v, and the size label decreases by a factor of $(1 - 2^{-\alpha(A)})$ when going to the right child.
(3) The leaves of the tree are the nodes for which either rank or size is equal to 1. These correspond to the submatrices in our partition.

To evaluate the number of leaves in the tree, match each leaf except the left most one to its closest ancestor where the rank label decreased for the last time. A simple estimate gives that the number of leaves matched to nodes with rank label roughly $r/2^i$ is at most $(2^{\alpha(A)} \log mn)^i$. Assuming, without loss of generality, that the matrix contains no repeated rows or columns, we have that $mn \leq M^{2r(A)}$. Summing up over i gives that the number of leaves is at most

$$2^{O(\log r(A)(\alpha(A)+\log r(A)+\log\log M))},$$

for every integer matrix A containing at most M different integer values. ∎

Using almost the same proof, it is possible to derive alternative bounds on $p(A)$. Here is an example, restricted to boolean matrices. For a matrix A let $T(A)$ be the maximum k such that A contains a $k \times k$ lower triangular submatrix with 1's on the main diagonal and 0's above it. Also let $\alpha^0(A)$ be the minimum integer such that every $s \times t$ submatrix of A has a submatrix of rank 0 and size at least $2^{-\alpha(A)}st$. Then

Theorem 3.2: *For every $m \times n$ boolean matrix A it holds that*

$$\log p(A) = O(\log T(A)(\alpha^0(A) + \log\log mn)).$$

Note that $T(A) \leq r(A)$ and even $T(A) \leq \mathrm{fs}(A)$ for every sign matrix A. The above bound therefore generalizes the bound in Theorem 3.1. The presence of $T(A), r(A)$ or some other related measure like $\mathrm{fs}(A)$ is unavoidable here. Consider the identity matrix $\mathbf{I}_n$ for example, it has $p(\mathbf{I}_n) = n$ while $\alpha^0(\mathbf{I}_n) = \Theta(1)$. The right-hand side of the inequality in 3.2 must therefore include an additional complexity measure that attains a high value on the identity matrix.

Another corollary of Theorem 3.1 is that the log-rank conjecture is true if and only if $\alpha(A)$ is bounded from above by a polynomial in $\log(r(A))$. As observed by Nisan and Wigderson [37], it follows that the log-rank conjecture is true if and only if the following conjecture holds:

Conjecture 3.3: *There is a constant c such that every $m \times n$ boolean matrix A contains a submatrix of rank at most 1 and size at least*

$$\frac{mn}{2^{\log^c r(A)}}.$$

The above reformulation nicely highlights the difficulty in resolving the log-rank conjecture. It requires proving that when the rank of a boolean matrix is small, it contains a large rank-0 or rank-1 submatrix. Either proving or disproving such a statement is a challenging task that requires ingenuity.

To prove the log-rank conjecture one would need to relate the rank of a matrix A, a global property, to the maximum size of a rank-1 submatrix, a local property. These kind of relations are known to be hard to prove. Coming up with a counterexample to the log-rank conjecture requires finding an explicit boolean matrix A with low rank that does not contain large sub-matrices of small rank. In other words, it requires finding strong two-source extractors (alternatively, explicit bounds on bipartite Ramsey numbers), with the additional condition that the rank of the matrix is low. The problem of finding strong two-source extractors (or bounding bipartite Ramsey numbers) is hard as it is, and has found good solutions only in recent years. Adding the requirement that the overall rank is low makes the problem even harder.

3.2. An improved bound in terms of rank.

Even though the bound in Theorem 3.1 has not helped resolve the log-rank conjecture yet, it has given rise to a few improvements on the trivial bound $p(A) \leq 2^{r(A)}$. We describe these improvements in this and the following subsections. These improved bounds work by finding a better upper bound on $\alpha(A)$ in terms of rank (this subsection) or one of its variants (the next subsection), and then applying Theorem 3.1.

This subsection is devoted to proving the following result:

THEOREM 3.4 (Lovett [35, Theorem 1.1]): *For every boolean matrix A*

$$p(A) = 2^{\tilde{O}(\sqrt{r(A)})}.$$

NOTATION AND BASIC DEFINITIONS. We begin with some basic definitions and notations. It will be convenient for us in proving Theorem 3.4 to work with sign matrices, instead of boolean matrices. A **sign matrix** is a matrix whose entries are either 1 or -1. Note that both rank and p do not change much when we go from boolean values to the sign version. With a slight abuse of notation we think of a submatrix of a sign matrix in two ways. One is the conventional way and the other is as follows. Let A be an $m \times n$ matrix and B an $s \times t$ submatrix of A. Let $a_{i,j}$ be the (i,j)-th entry of A and similarly $b_{i,j}$ denotes an entry of B. We sometimes think of B, for purposes of computations alone, as an $m \times n$ matrix by padding it with zeros. This way we can compute the inner product of B and matrices with the same dimensions as A. For instance we have $\sum_{i,j} a_{i,j}b_{i,j} = \sum_{i,j} b_{i,j}b_{i,j} = \|B\|_2^2 = st$.

Define a **weight matrix** as a real matrix with non-negative entries that sum up to 1. For an $m \times n$ sign matrix A, an $m \times n$ weight matrix W, a submatrix B of A, and a value $v \in \{\pm 1\}$, let

$$W(B) = \sum_{i,j:b_{i,j}\neq 0} w_{i,j},$$

$$W(v,B) = \sum_{i,j:b_{i,j}=v} w_{i,j},$$

$$S(W,B) = \left|\sum_{i,j} w_{i,j}b_{i,j}\right|.$$

DISCREPANCY AND CORRUPTION. For the proof we need two relaxed versions of $\alpha(A)$, discrepancy and corruption. For corruption, instead of looking for a large rank-1 matrix, we search for a large matrix that is close to being low rank, a "corrupted" low rank matrix. Another option is to consider submatrices with large discrepancy between the two values. We define two notions of corruption, each with its own useful properties.

For the first definition, let A be an $m \times n$ sign matrix, let $0 \leq \rho < 1$, and let W be a weight matrix of the same dimensions as A. Let $f_W(A) \in \{1, -1\}$ be the value with less weight in A (that is, $W(f_W(A), A) \leq W(-f_W(A), A)$), breaking ties arbitrarily. Denote by $\alpha_\rho(W, A)$ the minimum integer k such that there exists a submatrix B of A satisfying $W(B) \geq 2^{-k}$ and $W(f_W(A), B)/W(B) \leq \rho$. That is, the weight of B is at least 2^{-k}, and the weight of $f_W(A)$'s in B is at most ρ times the weight of B.

The second definition is a variant of α_ρ, hence we use the same notations as the previous paragraph. Denote by $\alpha^*_\rho(W, A)$ the minimum integer k such that A contains a submatrix B satisfying

$$W(B) \geq 2^{-k} \quad \text{and} \quad W(f_W(A), B)/W(B) \leq \rho W(f_W(A), A).$$

Note that $\alpha(A) \leq \alpha_0(U, A) = \alpha^*_0(U, A)$, where $U = \frac{1}{mn} J$ is the matrix with uniform weights (recall that J is the all-ones matrix).

Finally, we define discrepancy, which takes a complementary point of view. In discrepancy the focus is on the maximum possible discrepancy between the weight of entries of each type, instead of the weight of the submatrix. Let A be an $m \times n$ sign matrix, and let W be a weight matrix of the same dimensions as A. The discrepancy of A with respect to W, denoted by $\mathrm{disc}(W, A)$, is the minimum value of $S(W, B)^{-1}$ over all sub-matrices B of A. Note that here the weight of B is not taken into account. The discrepancy of A, denoted $\mathrm{disc}(A)$, is the maximum value of $\mathrm{disc}(W, A)$ over all weight matrices W of the same dimensions as A.

The plan. A core theme of [11, 16, 35], leading to the improved bound in terms of rank, is to replace $\alpha(A)$ with discrepancy somehow. This tempting approach was raised already by Nisan and Wigderson in [37]. They proved that $\mathrm{disc}(U, A) \leq r^{3/2}(A)$, for every sign matrix A. This bound was improved to $\mathrm{disc}(A) \leq \sqrt{r(A)}$; see [29, 30] for details. It is therefore left to prove an upper bound on communication complexity in terms of discrepancy. Such bounds were known in the literature, but the best one gave $cc(A) \leq \mathrm{disc}^2(A)$ which is insufficient for our purposes, as this way we can only retrieve the trivial bound $p(A) \leq 2^{r(A)}$ again. A main contribution of [11, 16, 35] is therefore an improved bound in term of discrepancy, and a main tool is amplification, that is, starting with a submatrix with a slight bias towards one of the signs (say $+1$) and amplifying it until we reach a submatrix with no -1 entries. We therefore break the proof into four steps. The first three implement the amplification process, and the last one employs the relation between discrepancy and rank. Note that the amplification process is where we use the definitions of corruption, to bridge between discrepancy and $\alpha(A)$, since the definitions of corruption (especially α^*_ρ) naturally allows amplification.

The steps of the proof are:

(1) Relate $\mathrm{disc}(W, A)$ to $\alpha_\rho(W, A)$, for every weight matrix W. The ρ we get here though is close to $1/2$, $\rho = 1/2 - 1/4d$.
(2) Relate $\alpha_\rho(\Lambda, A)$ to $\alpha^*_\rho(U, A)$. Here Λ is a specific weight matrix defined later.
(3) Relate $\alpha^*_\rho(U, A)$ to $\alpha(A)$. This step uses the natural amplification properties of α^*_ρ to go from $\rho = 1/2 - 1/4d$ to arbitrarily small ρ.
(4) Use the fact that $\mathrm{disc}(A) \leq \sqrt{r(A)}$, for every sign matrix A.

3.2.1. *Step 1.* We show that it is possible to translate low discrepancy to the existence of a submatrix of (relatively) large weight having some bias between signs.

Lemma 3.5 ([35, 45]): *Let A be a sign matrix, let W be a weight matrix with the same dimensions as A, and let $d = \mathrm{disc}(W, A)$. Then*

$$\alpha_{1/2-1/4d}(W, A) \leq 1 + \log d.$$

Proof. By definition of $\mathrm{disc}(W, A)$ it follows that there is some submatrix B of A such that

$$S(W, B) = \left| \sum_{i,j} w_{i,j} b_{i,j} \right| \geq \frac{1}{d}. \tag{2}$$

First observe that, incurring a factor of 2, we can assume that $f_W(A) = f_W(B)$. That is, we can assume that in both A and B the same value has less weight. Otherwise, assume without loss of generality that $f_W(A) = -1$ but

$$\sum_{i,j} w_{i,j} b_{i,j} < 0.$$

Then the weight of entries outside of B is at least $\frac{1}{d}$. These entries can be partitioned into two sub-matrices, hence there is a submatrix B' of A for which

$$\sum_{i,j} w_{i,j} b'_{i,j} \geq \frac{1}{2d}. \tag{3}$$

Therefore, let B be a submatrix of A satisfying $f_W(A) = f_W(B)$ and

$$\left|\sum_{i,j} w_{i,j} b_{i,j}\right| \geq \frac{1}{2d}.$$

Denote by $P(B)$ (respectively $N(B)$) the weight of 1's (respectively -1's) in B. That is, $P(B) = W(1, B)$ and $N(B) = W(-1, B)$. We have

$$\sum_{i,j} w_{i,j} b_{i,j} = P(B) - N(B),$$

and

$$W(B) = P(B) + N(B).$$

Hence, in case $f_W(A) = f_W(B) = -1$ and $\sum_{i,j} w_{i,j} b_{i,j} \geq 0$ we get

$$\begin{aligned} N(B) &= \frac{1}{2}W(B) - \frac{1}{2}\sum_{i,j} w_{i,j} b_{i,j} \\ &\leq \frac{1}{2}W(B) - \frac{1}{4d} \\ &\leq \frac{1}{2}W(B) - \frac{1}{4d}W(B) \\ &\leq \Big(\frac{1}{2} - \frac{1}{4d}\Big)W(B). \end{aligned}$$

In the other possible case, that $f_W(A) = f_W(B) = 1$ and $\sum_{i,j} w_{i,j} b_{i,j} \leq 0$, we similarly have

$$\begin{aligned} P(B) &= \frac{1}{2}W(B) + \frac{1}{2}\sum_{i,j} w_{i,j} b_{i,j} \\ &\leq \frac{1}{2}W(B) - \frac{1}{4d} \\ &\leq \frac{1}{2}W(B) - \frac{1}{4d}W(B) \\ &\leq \Big(\frac{1}{2} - \frac{1}{4d}\Big)W(B). \end{aligned}$$

In both cases, we find that the sign with less weight in A has weight at most $(\frac{1}{2} - \frac{1}{4d})W(B)$ in B, which concludes the requirements as

$$W(B) \geq \left|\sum_{i,j} w_{i,j} b_{i,j}\right| \geq \frac{1}{2d}.$$

We conclude that $\alpha_{1/2-1/4d}(A) \leq 1 + \log d$. ∎

3.2.2. *Step 2.* Let A be a sign matrix, and denote $\Lambda = \Lambda_A$ the weight matrix defined by

$$a_{i,j} = -1 \Rightarrow \lambda_{i,j} = \frac{1}{2N(A)}, \tag{4}$$

$$a_{i,j} = 1 \quad \Rightarrow \lambda_{i,j} = \frac{1}{2P(A)}. \tag{5}$$

Lemma 3.6: *For every sign matrix A and $0 \leq \rho \leq 1/2$, we have that*

$$\alpha^*_{2\rho}(U, A) \leq \alpha_\rho(\Lambda, A).$$

The proof of this lemma is straightforward, but technical. If the reader is familiar with the idea of boosting, then in a way this uses a similar idea. The intuition is revealed when considering the case of a balanced sign matrix A, satisfying $P(A) = N(A)$. In this case, the two notions $\alpha^*_{2\rho}(U, A)$ and $\alpha_\rho(\Lambda, A)$ exactly coincide. We leave it to the reader to complete the details of the general case, or look at the proof of Lemma 7 in [45].

3.2.3. *Step 3.* So far we have that

$$\alpha(A) \leq \alpha_0(U, A) = \alpha^*_0(U, A)$$

and

$$\alpha^*_{1-1/2d}(U, A) \leq \alpha_{1/2-1/4d}(\Lambda, A) \leq 1 + \log d,$$

where $d = \mathrm{disc}(A)$. It remains to relate $\alpha^*_0(U, A)$ and $\alpha^*_{1-1/2d}(U, A)$. The following lemmas give a tool to amplify the bias and go from $1-1/2d$ to arbitrarily small bias.

Lemma 3.7: *For every $m \times n$ sign matrix A, and every $0 \leq \rho_1, \rho_2 \leq 1$, there exists a weight matrix W such that*

$$\alpha^*_{\rho_1\rho_2}(U, A) \leq \alpha^*_{\rho_1}(U, A) + \alpha^*_{\rho_2}(W, A),$$

where W is uniform over some submatrix B of A.

Proof. Let $k_1 = \alpha^*_{\rho_1}(U, A)$, then by the definition of $\alpha^*_{\rho_1}$ there exists an $s \times t$ submatrix B of A satisfying $U(B) \geq 2^{-k_1}$ and $U(f_U(A), B)/U(B) \leq \rho_1 U(f_U(A), A)$. For simplicity assume without loss of generality that $f_U(A) = -1$, then we have

$$U(-1, B) \leq \rho_1 U(-1, A)\frac{st}{mn}.$$

Let W be the uniform weight matrix over the entries of B. That is, for entries outside B the weight is 0, and the weight of entries in B is $1/st$. Note that the above inequality implies $f_W(A) = f_U(A) = -1$. Let $k_2 = \alpha^*_{\rho_2}(W, A)$, then there exists a submatrix C of B satisfying

$$W(C) \geq 2^{-k_2} \quad \text{and} \quad W(-1, C)/W(C) \leq \rho_2 W(-1, A).$$

Therefore

$$\begin{aligned} U(-1, C)/U(C) &= W(-1, C)/W(C) \\ &\leq \rho_2 W(-1, A) \\ &= \rho_2 W(-1, B) \\ &= \rho_2 U(-1, B)\frac{mn}{st} \\ &\leq \rho_2 \rho_1 U(-1, A)\frac{st}{mn}\frac{mn}{st} \\ &\leq \rho_2 \rho_1 U(-1, A). \end{aligned}$$

To conclude the proof observe that

$$U(C) = \frac{|C|}{mn} = \frac{st}{mn}\frac{|C|}{st} = U(B)W(C) \geq 2^{-k_1}2^{-k_2}. \qquad \blacksquare$$

Using Lemma 3.7 repeatedly gives:

Corollary 3.8: *For every sign matrix A, $0 \leq \rho \leq 1$, and an integer $l > 0$, there exist weight matrices $W_1, \ldots, W_l$ such that*

$$\alpha^*_{\rho^l}(U, A) \leq \sum_{i=1}^{l} \alpha^*_\rho(W_i, A),$$

where each W_i is uniform over some submatrix B_i of A.

Lemma 3.7 and Corollary 3.8 allow to amplify the bias, which is almost all we need; we also need a way of stopping this process at some point. This breaking point is given by the following lemma, which is a nice contribution of [11, 16, 35].

Lemma 3.9 ([16]): *Let A be an $m \times n$ sign matrix with $r = r(A)$. Assume that either the fraction of 1's or the fraction of -1's in A is at most $1/4r$. Then A contains a rank-1 submatrix of size at least $\frac{mn}{4}$.*

Combining everything together. Let A be a sign matrix with $r = r(A)$. Lemma 3.9 implies that

$$\alpha(A) \leq \alpha_{1/4r}(U, A) + 2 \leq \alpha^*_{1/4r}(U, A) + 2.$$

Let $l = 2d(2 + \lceil \log r \rceil)$, then $(1 - 1/2d)^l \leq 1/4r$. Therefore, by Corollary 3.8, there exist weight matrices $W_1, \ldots, W_l$ such that

$$\alpha^*_{\rho^l}(U, A) \leq \sum_{i=1}^{l} \alpha^*_{\rho}(W_i, A).$$

Since, as stated in Corollary 3.8, each W_i is uniform over some submatrix B_i of A, we get from Lemma 3.6 that

$$\alpha^*_{1-1/2d}(W_i, A) \leq \alpha_{1/2-1/4d}(\Lambda_i, A).$$

Applying Lemma 3.5 we conclude that

$$\alpha(A) = O(d \log d \log r).$$

3.2.4. *Step 4.* It is left to use the fact that $\text{disc}(A) \leq \sqrt{r(A)}$, and to apply Theorem 3.1. The fact that $\text{disc}(A) \leq \sqrt{r(A)}$ follows from the John Ellipsoid Theorem [19], though to see that it might be necessary to first understand the relation between the discrepancy of a sign matrix and the factorization norm γ_2. The details can be found in [29, 30].

3.3. Approximate rank and sign rank.

As mentioned, it is a simple exercise to prove $p(A) \leq 2^{r(A)}$. In the last section we saw that this can be improved to $p(A) = 2^{\tilde{O}(\sqrt{r(A)})}$, but this is still very far from what the log-rank conjecture asserts. In this section we show that the trivial upper bound can be improved in a different way, replacing rank with weaker (and even much weaker) variants of it.

3.3.1. *Approximate rank.* Let A be a sign matrix and $0 < \epsilon < 1$ a real number. The **ϵ-approximate rank** of A is defined as

$$r_\epsilon(A) := \min_{B : \|A-B\|_\infty \leq \epsilon} r(B). \tag{6}$$

The approximate rank of a matrix can be significantly smaller than its rank. The $n \times n$ identity matrix $\mathbf{I}_n$ is a good example where rank is full but approximate rank (for constant ϵ) is logarithmic. The upper bound on $r_\epsilon(\mathbf{I}_n)$ follows from the Johnson–Lindenstrauss lemma [20], and a lower bound is given in [5]. More details and further examples can be found in [26], where it is shown that approximate rank behaves much more like the γ_2 factorization norm than like rank.

Despite the big difference between rank and approximate rank, we have

Lemma 3.10: *For every sign matrix A and $0 < \epsilon < 1$ it holds that*

$$p(A) = 2^{O(r_\epsilon(A))}.$$

Proof. The proof works via the notion of a cover number of an $m \times n$ sign matrix A. Denoted $N_\epsilon(A)$, it is the cover number of the convex hull of the columns of A. That is, it is the minimal size of a subset $S \subset \mathbb{R}^m$ such that for every vector v in the convex hull of the columns of A there is a vector $u \in S$ satisfying $\|v - u\|_\infty \leq \epsilon$.

It is proved in [6] that $N_\epsilon(A) \leq 2^{O(r_\epsilon(A))}$ (this inequality holds for every matrix whose entries are bounded by 1 in absolute value). It follows straight from the definitions that the number of distinct columns of A is bounded from above by $N_\epsilon(A)$, for every $0 < \epsilon < 1$. It is left to observe that $p(A)$ is at most the number of distinct columns of A. ∎

Note that the proof of Lemma 3.10 gives in fact that the number of distinct rows of a sign matrix A is at most $2^{O(r_\epsilon(A))}$, which is a stronger bound, as the number of distinct rows can be much larger than $p(A)$. In fact, for the number of distinct rows, the bound in terms of rank can be improved to $O(2^{r(A)/2})$, but no further; see [23] for details. Hence the bound in Lemma 3.10 in terms of approximate rank is a nice improvement.

3.3.2. *Sign rank.* Rank can be further replaced by **sign-rank** in terms of its relation with the maximum size of a rank-1 submatrix. The sign-rank of a sign matrix A is

$$r^\infty(A) := \min_{B: a_{i,j} b_{i,j} > 0} r(B).$$

Sign-rank is an even weaker version of rank than approximate rank, however, it still holds that:

Lemma 3.11: *For every $m \times n$ sign matrix A, $\alpha(A) = O(r^\infty(A))$.*

We use the following definition and result from [36]. For a set of points P in $\mathbb{R}^d$ and a hyperplane H, we say that H **crosses** P if not all of the points in P lie in the same half-space defined by H.

THEOREM 3.12 (Matoušek's Partition Theorem): *Let P be a subset of m points in $\mathbb{R}^d$, and let t be an integer. Then P can be partitioned into t subsets P_i each containing $\Theta(m/t)$ points from P such that every hyperplane crosses at most $O(t^{1-1/d})$ subsets P_i.*

Proof of Lemma 3.11. Let $d = r^\infty(A)$, and let $\{x_i\}_{i=1}^m, \{y_j\}_{j=1}^n \subset \mathbb{R}^d$ be row vectors satisfying $a_{i,j}(x_i y_j^t) > 0$. Apply Theorem 3.12 with $P = \{x_i\}_{i=1}^m$ and $t = 2^{O(d)}$. We get a partition of $\{x_i\}_{i=1}^m$ into t subsets P_i each containing $\Theta(m/t)$ points such that the sign of the inner product of every y_j is constant on at least one of the subsets. This gives a monochromatic submatrix of A of size at least $mn/2^{O(d)}$. Since the sign-rank of any submatrix of A is at most the sign-rank of A, the lemma follows. ∎

4. Known gaps

In this section we review some of the best known separations between communication complexity and log-rank. For this section it will often be more convenient to think in terms of functions rather than matrices. For the separations we discuss a key role is played by the composition of functions and, while this can be described purely in matrix terms, it is cleaner to state using a functional notation.

We call a function of the form $F : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ a **communication function** as it describes the communication problem where Alice is given $x \in \{0,1\}^n$, Bob is given $y \in \{0,1\}^m$ and their goal is to communicate as little as possible until they both know $F(x,y)$. From a communication function F we can define a matrix A where $A_{x,y} = F(x,y)$. This is known as the **communication matrix** of F.

4.1. Nisan, Wigderson and Kushilevitz.

In this section we go over the construction by Nisan and Wigderson [37] of a boolean matrix A for which $cc(A) = \Omega(\log^{\log_2 3} r(A))$, implying that the constant c in the log-rank conjecture must be at least $\log_2 3 \approx 1.58$, and its improvement by Kushilevitz to show that $c \geq \log_3 6 = 1.63\ldots$.

This construction, and in fact all the separations between rank measures and communication complexity measures we will look at, is based on a composed function. A composed function is a way to transform a function from the "query" setting, where separations are usually easier to prove, to the "communication" or matrix setting.

Say that we have a function $f : \{0,1\}^n \to \{0,1\}$. On input x the goal of a deterministic query algorithm for f is to output $f(x)$, but the algorithm can only access x by making queries of the form: what is x_i? These questions can be made in an adaptive fashion, i.e., the choice of what to query can depend on the answer to previous queries. Let $D(f)$ be the minimum number of queries needed for a deterministic algorithm to correctly output $f(x)$ for all $x \in \{0,1\}^n$. As an example, consider the function $\mathrm{OR}_n : \{0,1\}^n \to \{0,1\}$ which evaluates to 0 if and only if the input is the all-zero string. It is easy to see that $D(\mathrm{OR}_n) = n$ by an **adversary** argument. For a query "what is x_i?" the adversary decides how to answer. In the case of the OR_n function the behavior of the adversary is very simple: for any query the algorithm makes, the adversary answers that the corresponding bit of the input is 0. After $n-1$ queries there is still some position of the input x which has not been queried. By deciding whether or not this last input bit is 0 or 1 the adversary determines whether $\mathrm{OR}_n(x) = 0$ or $\mathrm{OR}_n(x) = 1$. Thus the algorithm cannot correctly answer after $n-1$ queries, showing $D(\mathrm{OR}_n) \geq n$. Of course, $D(\mathrm{OR}_n) \leq n$ because with n queries the algorithm can learn the entire input. The property that makes the lower bound argument work is that the OR function has **sensitivity** n on the input 0^n. Flipping any bit of 0^n changes the output value under OR_n. Note that the adversary only constructs inputs that have at most one 1, thus the same argument also works to show that the UNIQUE OR function, which is only defined on inputs of Hamming weight at most 1, has deterministic query complexity n.

We can transform the function f from the "query world" into the "communication world" by composing it with a "gadget." Perhaps the simplest gadget is the function $\mathrm{AND} : \{0,1\} \times \{0,1\} \to \{0,1\}$ where $\mathrm{AND}(x,y) = 1$ if and only if $x = y = 1$. If we have a "query" function $f : \{0,1\}^n \to \{0,1\}$ then we can create a "communication" function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ as

$$F(x,y) = f(\mathrm{AND}(x_1,y_1),\ldots,\mathrm{AND}(x_n,y_n)).$$

We denote this function as $F = f \circ \mathrm{AND}^n$.

As an example, when $f = \mathrm{OR}_3$ then the communication matrix of $\mathrm{OR}_3 \circ \mathrm{AND}^3$ looks as follows:

	000	001	010	011	100	101	110	111
000	0	0	0	0	0	0	0	0
001	0	1	0	1	0	1	0	1
010	0	0	1	1	0	0	1	1
011	0	1	1	1	0	1	1	1
100	0	0	0	0	1	1	1	1
101	0	1	0	1	1	1	1	1
110	0	0	1	1	1	1	1	1
111	0	1	1	1	1	1	1	1

It is not hard to see that the rank of this matrix is 7. The function $\mathrm{OR}_n \circ \mathrm{AND}^n$ is known as the SET INTERSECTION function as it evaluates to 1 if and only if the inputs x, y share a common 1 in some position. This is one of the most important functions in communication complexity.

Nisan and Wigderson make one change to this matrix. If we change $(8, 8)$ entry from 1 to 0 then the rank of the matrix becomes 6. This again corresponds to a composed function, $\mathrm{NAE} \circ \mathrm{AND}^3$ where NAE is the "not-all-equal" function. This is identical to OR_3 except that $\mathrm{NAE}(111) = 0$. Thus NAE evaluates to 1 exactly when not all the input bits are equal. For the separating example Nisan and Wigderson compose NAE with itself k times to obtain a function $\mathrm{NAE}^k : \{0,1\}^{3^k} \to \{0,1\}$, then look at the function $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$. Let's see why this function gives a separation between log-rank and communication complexity.

The NAE function still has sensitivity 3 on the input 000. This property carries over when we compose NAE with itself—NAE^k has sensitivity 3^k on the input 0^{3^k}. This means that $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$ is an instance of the UNIQUE SET INTERSECTION— whenever x and y intersect in at most one position $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$ has the same output as SET INTERSECTION. It is known that UNIQUE SET INTERSECTION on m bit input has deterministic communication complexity $\Omega(m)$ [21, 41]. This shows the communication complexity lower bound $\Omega(3^k)$ on $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$.

The key to the upper bound on the rank of the communication matrix of $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$ is to look at the polynomial representation of NAE. The matrix rank of $f \circ \mathrm{AND}^n$ is closely related to the number of monomials in the polynomial representation of f. Suppose that $f(x) = \sum_{S\subseteq[n]} \alpha_S x_S$, where $x_S = \prod_{i\in S} x_i$. Then

$$f \circ \mathrm{AND}^n(x,y) = \sum_{S\subseteq[n]} \alpha_S x_S y_S$$

immediately gives a factorization of the communication matrix of $f \circ \mathrm{AND}^n$ that witnesses the rank is at most the number of monomials in f.

The not-all-equal function has polynomial degree 2:

$$\mathrm{NAE}(x_1x_2x_3) = 1 - x_1x_2 - x_1x_3 - x_2x_3.$$

Composing this polynomial with itself k times gives a polynomial representation of NAE^k, showing that $\deg(\mathrm{NAE}^k) \le 2^k$. Thus the number of monomials in NAE^k is at most $\binom{3^k}{2^k} \le 3^{k2^k}$. Therefore, the logarithm of the rank of the communication matrix of $\mathrm{NAE}^k \circ \mathrm{AND}^{3^k}$ is $O(k2^k)$. This gives an example of a matrix with communication complexity $\Omega(3^k)$ and logarithm of the rank $O(k2^k)$, showing that the exponent in the log-rank conjecture must be at least $\log_2 3 = 1.58\ldots$.

The essence of the Nisan–Wigderson construction is finding a separation between the polynomial degree of f and the sensitivity of f on input 0^n. By the connection to the UNIQUE SET INTERSECTION problem, the sensitivity of f on 0^n will always be a lower bound on the communication complexity of $f \circ \mathrm{AND}^n$. And $\binom{n}{\deg(f)}$ will always be an upper bound on the rank of $f \circ \mathrm{AND}^n$.

Kushilevitz improved on the Nisan–Wigderson example by coming up with a function f on 6 bits that has sensitivity 6 on input 0^6 and has polynomial degree 3. This example shows that the exponent in the log-rank conjecture must be at least $\log_3 6 = 1.63\ldots$

It remains an open problem how large the separation between polynomial degree and sensitivity can be. It is known that $s(f) \le \deg(f)^2$ and the example of Kushilevitz shows the existence of an infinite family of functions f with $\deg(f)^{\log_3 6} \le s(f)$. As far as we currently know, the Nisan–Wigderson framework could be instantiated to show that the exponent in the log-rank conjecture must be as large as 2. Such a separation was achieved by Göös, Pitassi and Watson [17] by very different means, as we see in the next section.

Question 4.1: Does there exist a function f whose sensitivity is quadratically larger than its degree?

4.2. Göös, Pitassi and Watson. Göös, Pitassi and Watson (GPW) [17] show that the exponent in the log-rank conjecture must be at least 2 by different means than those discussed in Section 4.1. Specifically, they give an example of a matrix A with $cc(A) = \tilde{\Omega}(\log^2 p(A)))$. This separation is tight, up to logarithmic factors, in light of Eq. (1).

The GPW separation again begins by showing a separation in the query setting and then "lifting" this separation to the communication setting. This time, instead of working with the AND function gadget, GPW use the **index function** INDEX_m as the gadget. This is defined as $\text{INDEX}_m : [m] \times \{0,1\}^m \to \{0,1\}$, where $\text{INDEX}_m(i,x) = x_i$. Theorem 3 of [17] shows a very powerful lifting theorem. For any $f : \{0,1\}^n \to \{0,1\}$, by taking $m = n^{20}$ this theorem shows that $cc(f \circ g^n) = D(f) \cdot \Theta(\log n)$. Recall that $D(f)$ is the deterministic query complexity of f. This theorem is powerful because it is generally much easier to show a lower bound on query complexity rather than communication complexity. Similar to what we saw in Section 4.1 the problem now becomes to construct a query function that has a separation between its polynomial degree and deterministic query complexity (upper bounds on degree similarly lift to upper bounds on rank with the INDEX_m gadget).

In the query setting, it is generally easier to separate complexity measures on partial functions, functions whose domain is not the whole boolean cube $\{0,1\}^n$, rather than total functions. For example, the polynomial $x_1 + \cdots + x_n$ has degree 1. When restricted to inputs that have at most one 1, this polynomial faithfully represents the UNIQUE OR promise problem. As we discussed in Section 4.1 the deterministic query complexity of UNIQUE OR is n. Thus we have a separation of 1 vs. n between polynomial degree and deterministic query complexity for the partial function UNIQUE OR. A remarkable insight of GPW is a technique for constructing separations by making partial functions total. This insight has been highly influential in later work, leading to the falsification of the Saks–Wigderson conjecture [9] and the largest known separations between randomized and quantum query complexities [1].

To explain the GPW function, we start with a partial function slightly more involved than UNIQUE OR. Now think of the input variables $x_{i,j}$ laid out in an n-by-n matrix. The function now is UNIQUE OR of AND: we want to determine if there is a column where all variables are 1, and are promised that there is at most one such column. There is a degree n polynomial for this partial function: $\sum_i \prod_j x_{i,j}$. On the other hand the deterministic query complexity of this function is $\Omega(n^2)$: an adversary can always answer a query 1 unless it is the last unknown variable in a column, in which case it answers 0. Now we have a quadratic separation between polynomial degree and deterministic query complexity on a partial function.

The advantage of the UNIQUE OR of AND function, and a key insight of GPW, is that it can be made total without changing the asymptotic separation. This is done via the use of **pointers**. Each cell of the n-by-n matrix is now additionally given another piece of data, which is a pointer, possibly null (denoted $\perp$), to another cell of the matrix.

The function is defined to evaluate to 1 if and only if:

- There is a unique all-1 column.
- In the all-1 column there is a unique cell with a non-null pointer.
- Following this pointer starts a chain of pointers that visits a 0-cell of every other column.

Adding the verification of a unique all-1 column makes this a total function.

There is still a simple adversary argument to show that this function has query complexity $\Omega(n^2)$. The adversary always answers $(1, \perp)$ unless a query is the last unqueried cell in a column, a "critical query." On the first critical query the adversary answers $(0, \perp)$. On subsequent critical queries the adversary answers with 0 and a pointer to the cell where the previous critical query was made.

We can construct a $O(n)$ degree polynomial which encodes that a column is all-1 and starts a pointer chain to zeros in all other columns. As there can be at most one such column, we can sum these polynomials over all columns to give a degree $O(n)$ polynomial for the GPW function overall. Composing the GPW query function with the INDEX gadget gives a communication function F whose deterministic communication complexity is $\tilde{\Omega}(n^2)$ while the communication matrix A of F satisfies $\log p(A) = \tilde{O}(n)$.

5. Variations, including false ones

In this section we look at variations around the log-rank conjecture. By replacing rank with other complexity measures, like non-negative rank or approximate rank, we can obtain statements that are known to be true or false. We also examine the hypotheses of the log-rank conjecture, in particular the assumption that the matrix is boolean.

5.1. Integer matrices as opposed to boolean matrices. The definition of partition number is easily extended to non-negative integer matrices. For $A \in \mathbb{N}^{m \times n}$ we define $p(A)$ to be the smallest p such that there exist boolean matrices $\{B_i\}_{i=1}^p$ and positive integers $\{z_i\}_{i=1}^p$ such that $r(B_i) = 1$ and $A = \sum_{i=1}^p z_i B_i$.

In the introduction we mentioned that the $n \times n$ matrix A with $a_{i,j} = i + j$ has rank 2 but partition number n because no two diagonal entries of A can belong to the same part of the partition. A drawback to this example is that the entries of the matrix are rather large, as large as $2n$.

We can achieve a separation that is almost as good with an $N \times N$ matrix whose entries are at most $\log N$. Define an $N \times N$ matrix A where $N = 2^n$ and rows and columns are labeled by n-bit strings. For $x, y \in \{0,1\}^n$ define $a_{x,y} = \langle x, y \rangle$, where $\langle \cdot, \cdot \rangle$ is the dot product. This explicit factorization by n-dimensional vectors shows $r(A) \leq n$. Further, when restricted to entries (x, y) where $\langle x, y \rangle \leq 1$ the matrix A agrees with the UNIQUE SET INTERSECTION problem, which shows that $\log p(A) = \Omega(n)$.

Another source of interesting examples of non-negative matrices that have small rank, large partition numbers, and whose entries are not too large comes from looking at slack matrices of polytopes. A polytope $P \subseteq \mathbb{R}^d$ has two natural representations. It can be represented as the convex hull of its vertices $P = \mathrm{conv}(v_1, \ldots, v_t\}$ or it can be represented by its defining inequalities $P = \{x \in \mathbb{R}^d : Ax \leq b\}$, where the matrix A has f rows. The slack matrix M of P is a t-by-f matrix with rows labeled by vertices and columns labeled by defining inequalities where $m_{i,j} = b_j - A_j v_i$ is the **slack** of the i^{th} vertex with respect to the j^{th} inequality. In particular, the entries of a slack matrix are non-negative. This definition also shows that the rank of a slack matrix is at most $d + 1$ as, for V the matrix whose i^{th} row is v_i^T, it can be written as

$$M = \begin{bmatrix} \mathbf{1}_{t\times 1} & V \end{bmatrix} \begin{bmatrix} b^T \\ -A^T \end{bmatrix},$$

where $\mathbf{1}_{t\times 1}$ is a column vector of dimension t consisting of all ones.

Slack matrices can provide interesting examples of low rank matrices. We will be interested in the slack matrix of the correlation polytope. The vertices of the correlation polytope are aa^T for all $a \in \{0,1\}^n$. For any $a, b \in \{0,1\}^n$

we of course have $(1-a^Tb)^2 \geq 0$. We can rewrite this as

$$1 - 2\text{diag}(a)bb^T + \langle aa^T, bb^T\rangle.$$

This means that every vertex bb^T of the correlation polytope satisfies the inequality $\langle 2\text{diag}(a)-aa^T, bb^T\rangle \leq 1$. Hence the 2^n-by-2^n matrix $m_{a,b} = (1-a^Tb)^2$ is a submatrix of the slack matrix of the correlation polytope. This matrix has rank at most n^2+1 and all entries have magnitude at most $(n-1)^2$. This submatrix is again an instance of (the negation of) UNIQUE SET INTERSECTION, thus it satisfies $\log p(M) = \Omega(n)$. See [15] for more details about this example, where it is used to show an exponential lower bound on the extension complexity of the cut polytope.

5.2. The log-approximate-rank conjecture. By analogy with the log-rank conjecture, in 2009 we proposed the log approximate rank conjecture [27]. This conjecture replaces deterministic communication complexity with (private coin) randomized communication complexity, and rank with approximate rank. We let $\text{rcc}_{1/3}(A)$ denote the minimum cost of a private coin randomized communication protocol that correctly computes A. As the logarithm of the rank is a lower bound on deterministic communication complexity, the logarithm of the approximate rank is a lower bound on randomized communication complexity (recall the definition of approximate rank from Eq. (6)). Krause [24] showed that $\text{rcc}_{1/3}(A) \geq \log r_{1/3}(A)$. The log approximate rank conjecture is that there is a universal constant c such that $\text{rcc}_{1/3}(A) \leq \log(r_{1/3}(A))^c + 2$. This conjecture was recently shown to be false by Chattopadhyay, Mande and Sherif (CMS) [12]. Shortly thereafter, an even weaker version of the log-approximate-rank conjecture with randomized communication complexity replaced by quantum communication complexity was also refuted by two groups, Anshu, Boddu and Touchette [10], and Sinha and de Wolf [46].

As in the Nisan–Wigderson and Göös, Pitassi and Watson constructions, the counterexample is again based on achieving a separation in the query world and lifting this separation to communication complexity. The analog of randomized communication complexity in the query setting is randomized query complexity, which we denote by $R_{1/3}(f)$. The analog of approximate rank in the query setting is the minimum number of monomials in a polynomial approximating f.

To achieve a super-polynomial separation between randomized communication complexity and the logarithm of approximate rank we have to be careful about which kind of gadget g we choose. For a boolean function $f:\{0,1\}^n \to \{0,1\}$ let $\deg_{1/3}(f)$ be the **approximate degree** of f, that is the minimum degree of a real-valued function h such that $|f(x)-h(x)| \leq 1/3$ for all $x \in \{0,1\}^n$. Work by Sherstov on the pattern matrix method [42] and Shi and Zhu [43] shows for a variety of gadgets g that $\log r_{1/3}(f \circ g^n) = \Omega(\deg_{1/3}(f))$. One statement of this form by Lee and Zhang [28] says the following. Say that a row (column) of a sign matrix A is balanced if the sum of the entries in the row (column) is zero. Say that the sign matrix A is **strongly balanced** if every row and column is balanced. For $g : X \times Y \to \{0,1\}$ let M_g be a sign matrix with $M_g(x,y) = (-1)^{g(x,y)}$. For any function f, if g is such that M_g is strongly balanced it follows that

$$\log r_{1/3}(f \circ g^n) = \Omega\Big(\deg_{1/3}(f) \frac{\sqrt{|X||Y|}}{\|M_g\|}\Big). \tag{7}$$

We further know that $R_{1/3}(f) \leq \deg_{1/3}(f)^4$ [2], so we cannot have a superpolynomial separation between the log approximate rank and randomized communication complexity of a function of the form $f \circ g^n$ when g is strongly balanced and has $\sqrt{|X||Y|}/\|M_g\| > 1$.

A way around this is to use as the inner gadget the XOR function $g(x,y)=x \oplus y$ where $x, y \in \{0,1\}$. While this is a strongly balanced function, in this case M_g has rank one and therefore $\sqrt{|X||Y|}/\|M_g\| = 1$, so Eq. 7 gives nothing. Indeed, even though PARITY has polynomial degree n, the function $\mathrm{PARITY}_n \circ \mathrm{XOR}^n$ can be computed by a 2-bit communication protocol. When the gadget is the XOR function, the analogs of approximate rank and randomized communication complexity in the query world change. The rank of $f \circ \mathrm{XOR}^n$ is exactly $\|\hat{f}\|_0$, the number of nonzero Fourier coefficients of f, or in other words the number of monomials in a polynomial representing f as a function over the domain $\{-1,+1\}^n$. Likewise, the approximate rank of $f \circ \mathrm{XOR}^n$ is determined by $\|\hat{f}\|_{0,1/3} = \min_h\{\|\hat{h}\|_0 : \|f-h\|_\infty \leq 1/3\}$. Instead of the normal randomized query complexity, the randomized communication complexity of $f \circ \mathrm{XOR}^n$ is connected to a more powerful query model called a randomized parity decision tree. In this model, the algorithm can query the parity of any subset of the input bits. We use $R_\oplus(f)$ to denote the minimum cost of a randomized parity decision tree that computes f with error at most $1/3$. We can see that $\mathrm{rcc}_{1/3}(f \circ \mathrm{XOR}^n) = O(R_\oplus(f))$ because $\mathrm{PARITY}_k \circ \mathrm{XOR}^k$ can be computed with constant communication.

CMS consider the case of the SINK function $\mathrm{SINK}_m : \{0,1\}^{\binom{m}{2}} \to \{0,1\}$. The input $x \in \{0,1\}^{\binom{m}{2}}$ is interpreted as a directed graph on m vertices. In fact, x is interpreted as being a **tournament**: for every pair of vertices v_i, v_j there is either an edge directed from v_i to v_j or an edge directed from v_j to v_i, but not both. Indexing the bits of the input as $x_{i,j}$ for $i < j$ then $x_{i,j} = 1$ means there is an edge from v_i to v_j in the graph, and $x_{i,j} = 0$ means there is an edge from v_j to v_i in the graph. In the query world, CMS show the following separation:

(1) $\|\widehat{\mathrm{SINK}}_m\|_{0,1/3} = O(m^4)$.

(2) $R_\oplus(\mathrm{SINK}_m) = \Omega(m)$.

Item (1) directly implies that $r_{1/3}(\mathrm{SINK}_m \circ \mathrm{XOR}^{\binom{m}{2}}) = O(m^4)$. While we do not have a general theorem lifting randomized parity decision tree lower bounds to communication lower bounds for XOR functions, for the particular case of the SINK function, CMS are able to lift item (2) to show $\mathrm{rcc}_{1/3}(\mathrm{SINK} \circ \mathrm{XOR}^{\binom{m}{2}}) = \Omega(m)$. Thus this gives an example of an exponential separation between the logarithm of the approximate rank and randomized communication complexity.

5.3. Non-negative rank. There is a variation of the log-rank conjecture that we know to be true—if rank is replaced by the non-negative rank. The non-negative rank of an $m \times n$ non-negative matrix A, denoted $r_+(A)$, is the minimum r such that there is an $m \times r$ non-negative matrix X and an $n \times r$ non-negative matrix Y such that $A = XY^t$. We can see that $\chi(A) \le r_+(A)$. Indeed, say $r_+(A) = r$ and take a non-negative factorization $A = XY^t$ where X and Y have r columns. From X, Y we define r boolean rank-one matrices $B_1, \ldots, B_r$ where B_i is is 1 wherever $X_i Y_i^t$ is positive and is 0 otherwise. Since all entries of X, Y are non-negative there can be no cancelations, and A must be 1 in any entry where $X_i Y_i^t$ is positive. Moreover, as $A = XY^t$ it must be the case that for every entry where A is one, there is some B_i that is 1 there. Thus $B_1, \ldots, B_r$ form a covering of the ones of A and $\chi(A) \le r_+(A)$. Therefore by Lemma 5.1 below we have

$$\log p(A) \le (\log r_+(A) + 2)(\log r(A) + 1).$$

Lemma 5.1 (Lovász and Saks [34, Theorem 2.8]): *For every boolean matrix A it holds that*

$$\log p(A) \le (\log \chi(A) + 2)(\log r(A) + 1).$$

6. Conclusion

The log-rank conjecture is notorious for good reasons:

- It is natural and elementary, as it asks whether the basic measure in communication complexity, an information theoretic model, is characterized by the first proved lower bound, matrix rank. Nevertheless after more than 50 years of research, we are not even close to resolving this question.
- It has profound applications, since matrix rank has many properties we do not know whether communication complexity possesses:
  - It can be computed in polynomial time.
  - It has two dual equivalent characterizations.
  - The rank of a tensor product is equal to the product of the ranks.
  - For every matrix A, with $r = r(A)$, there is an $r \times r$ submatrix of rank r.
- Generalizing the log-rank conjecture a bit, for example by considering any integer matrix, results in a false statement. And making the statement stronger, e.g., by replacing rank by binary or positive rank, results in a (relatively) easy theorem. For this reason, and others, it seems to carve a thin line.
- Restricting the question to a small subset of matrices, e.g., matrices A such that

  $$A_{x,y} = f(x \oplus y)$$

  for some function $f\{0,1\}^n \to \{0,1\}$, is still a hard and interesting question. Even restricting and asking for the minimum rank of a triangle-free graph, still seems quite challenging.
- To improve the upper bound $\chi(A) \leq 2^{\tilde{O}(\sqrt{r(A)})}$ one needs to relate matrix rank, a global property, to the maximal size of a rank-one matrix, a local property. These kind of relations are known to be challenging.
- To improve the known gap between rank and χ, one needs to construct a boolean matrix A with low rank that does not contain large submatrices of rank 1. In other words, it requires finding strong two-source extractors (alternatively, explicit bounds on bipartite Ramsey numbers), with the additional condition that the rank of the matrix is low. The problem

of finding strong two-source extractors (or bounding bipartite Ramsey numbers) is hard as it is, and has found good solutions only in recent years; adding the requirement that the overall rank is low makes the problem even harder.

- If we instead deviate from this path, and attempt a non-explicit construction we would have to face the fact that we have no way of sampling low-rank boolean matrices. There is no random model for this set of matrices.

Even though the log-rank conjecture seems far from our reach at the moment, we did learn quite a bit from the reseach so far. For example, evidence is mounting via the Nisan and Wigderson bound and related work, e.g., [34, 16], that when rank is small things are strongly structured. When rank is small, many complexity measures that are usually very different from each other, become close. This includes: discrepancy, corruption, the maximal size of a rank-1 submatrix, deterministic communication complexity, non-deterministic communication complexity, randomized communication complexity, and many more. This even includes quantum communication complexity measures and a measure that combines the power of randomization and non-determinism. Perhaps the reason for this structure is that the rank of a boolean matrix can only be made small by repeating a row or a column, more or less. For a random matrix this is a commonly accepted conjecture, and it is nearly resolved [31]. If true, this can explain the strong structure implied by small rank, and it will also imply the log-rank conjecture. Or, it might be the case that there exists some extraordinary way to span many boolean vectors from some small set of vectors, that allow to refute the log-rank conjecture.

Either way, the log-rank conjecture highlights a hole in our understanding. It points to an intersection between linear algebra and combinatorics where our usual tools do not work. When restricted to boolean matrices, we do not understand rank well enough, and even more so restricted versions of rank, or quantities related to the size of submatrices with small rank. There are various threads to pool here, and different approaches one can take for the log-rank conjecture, each interesting in its own right. We hope that in this survey we managed to showcase some of these fascinating aspects of the log-rank conjecture.

References

[1] S. Aaronson, S. Ben-David and R. Kothari, *Separations in query complexity using cheat sheets*, in *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, ACM, New York, 2016, pp. 863–876.

[2] S. Aaronson, S. Ben-David, R. Kothari, S. Rao and A. Tal, *Degree vs. approximate degree and quantum implications of Huang's sensitivity theorem*, in *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, ACM, New York, 2021, pp. 1330–1342.

[3] A V. Aho, J. D. Ullman and M. Yannakakis, *On notions of information transfer in VLSI circuits*, in *STOC '83–Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, ACM, New York, 1983, pp. 133–139.

[4] M. Ajtai, J. Komlós and E. Szemerédi, *A note on Ramsey numbers*, Journal of Combinatorial Theory. Series A **29** (1980), 354–360.

[5] N. Alon, *Perturbed identity matrices have high rank: proof and applications*, Combinatorics, Probability, and Computing **18** (2009), 3–15.

[6] N. Alon, T. Lee, A. Shraibman and S. Vempala, *The approximate rank of a matrix and its algorithmic applications*, in *TOC'13—Proceedings of the 2013 ACM Symposium on Theory of Computing*, ACM, New York, 2013, pp. 675–684

[7] N. Alon and P. Seymour, *A counterexample to the rank-coloring conjecture*, Journal of Graph Theory **13** (1989), 523–525.

[8] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & sons, Hoboken, NJ, 2008.

[9] A. Ambainis, K. Balodis, A. Belovs, T. Lee, M. Santha and J. Smotrovs, *Separations in query complexity based on pointer functions*, Journal of the ACM **64** (2017), Article no. 32.

[10] A. Anshu, N. Boddu and D. Touchette, *Quantum log-approximate-rank conjecture is also false*, in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society, Los Alamitos, CA, 2019, pp. 982–994.

[11] E. Ben-Sasson, S. Lovett and N. Ron-Zewi, *An additive combinatorics approach relating rank to communication complexity*, in *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, IEEE Computer Society, Los Alamitos, CA, 2012, pp. 177–186.

[12] A. Chattopadhyay, N. S. Mande and S. Sherif, *The log-approximate-rank conjecture is false*, Journal of the ACM **67** (2020), Article no. 23.

[13] M. Dietzfelbinger, J. Hromkovič and G. Schnitger, *A comparison of two lower-bound methods for communication complexity*, Theoretical Computer Science **168** (1996), 39–51.

[14] S. Fajtlowicz, *On conjectures of Graffiti*, Discrete Mathematics **72** (1988), 113–118.

[15] S. Fiorini, S. Massar, S. Pokutta, H. Tiwary and R. de Wolf, *Exponential lower bounds for polytopes in combinatorial optimization*, Journal of the ACM **62** (2015), Article no. 17.

[16] D. Gavinsky and S. Lovett, *En route to the log-rank conjecture: New reductions and equivalent formulations*, in *Automata, Languages, and Programming* Lecture Notes in Computer Science, Vol. 8572, Springer, Heidelberg, 2014, pp. 514–524.

[17] M. Göös, T. Pitassi and T. Watson, *Deterministic communication vs. partition number*, SIAM Journal on Computing **47** (2018), 2435–2450.

[18] B. Halstenberg and R. Reischuk, *On different modes of communication*, in *STOC '88–Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, ACM, New York, 1988, pp. 162–172.

[19] F. John, *Extremum problems with inequalities as subsidiary conditions*, in *Traces and Emergence of Nonlinear Programming*, Birkhäuser/Springer, Basel, 2014, pp. 197–215

[20] W. Johnson and J. Lindenstrauss, *Extensions of Lipschitz mapping into Hilbert space*, in *Conference in modern analysis and probability (New Haven, Conn., 1982)*, Contemporary Mathematics Vol. 26, American Mathematical Society, providence, RI, 1984, pp. 189–206.

[21] B. Kalyanasundaram and G. Schnitger, *The probabilistic communication complexity of set intersection*, SIAM Journal on Discrete Mathematics **5** (1992), 545–557.

[22] A. Kotlov, *Rank and chromatic number of a graph*, Journal of Graph Theory **26** (1997), 1–8.

[23] A. Kotlov and L. Lovász, *The rank and size of graphs*, Journal of Graph Theory **23** (1996), 185–189.

[24] M. Krause, *Geometric arguments yield better bounds for threshold circuits and distributed computing*, Theoretical Computer Science **156** (1996), 99–117.

[25] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, Cambridge, 1997.

[26] T. Lee and A. Shraibman, *An approximation algorithm for approximation rank*, in *24th Annual IEEE Conference on Computational Complexity*, IEEE Computer Society, Los Alamitos, CA, pp. 351–357.

[27] T. Lee and A. Shraibman, *Lower bounds in communication complexity*, Foundations and Trends in Theoretical Computer Science **3** (2009), 263–398.

[28] T. Lee and S. Zhang, *Composition theorems in communication complexity*, in *Automata, Languages, and Programming*, Lecture Notes in Computer Science, Vol. 6198, Springer, Berlin, 2010, pp. 475–489.

[29] N. Linial and A. Shraibman, *Learning complexity versus communication complexity*, Combinatorics, Probability, and Computing **18** (2009), 227–245.

[30] N. Linial and A. Shraibman, *Lower bounds in communication complexity based on factorization norms*, Random Structures & Algorithms **34** (2009), 368–394.

[31] A. E. Litvak and K. E. Tikhomirov, *Singularity of sparse Bernoulli matrices*, Duke Mathematical Journal **171** (2022), 1135–1233.

[32] L. Lovász, *Communication complexity: A survey.* in *Paths, Flows, and VLSI-Layout*, Algorithms and Combinatorics, Vol. 9, Springer, berlin, 1990, pp. 235–265.

[33] L. Lovász and M. Saks, *Möbius functions and communication complexity*, in *29th IEEE Annual Symposium on Foundations of Computer Science—FOCS 1988*, IEEE Computer Society, Los Alamitos, CA, 1988, pp. 81–90.

[34] L. Lovász and M. Saks. *Communication complexity and combinatorial lattice theory*, Journal of Computer and System Sciences **47** (1993), 322–349.

[35] S. Lovett, *Communication is bounded by root of rank*, Journal of the ACM **63** (2016), Article no. 1.

[36] J. Matoušek, *Efficient partition trees*, Discrete & Computational Geometry **8** (1992), 315–334.

[37] N. Nisan and A. Wigderson, *On rank vs. communication complexity*, Combinatorica **15** (1995), 557–566.

[38] C. van Nuffelen, *A bound for the chromatic number of a graph*, American Mathematical Monthly **83** (1976), 265–266.

[39] R. Raz and B. Spieker, *On the "log-rank" conjecture in communication complexity*, Combinatorica **15** (1995), 567–588.

[40] A. Razborov, *The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear*, Discrete Mathematics **108** (1992), 393–396.

[41] A. Razborov, *On the distributional complexity of disjointness*, Theoretical Computer Science **106** (1992), 385–390.

[42] A. Sherstov, *The pattern matrix method*, SIAM Journal on Computing **40** 2011, 1969–2000.

[43] Y. Shi and Y. Zhu, *Quantum communication complexity of block-composed functions*, Quantum Information & Computation **9** (2009), 444–460.

[44] M. Shigeta and K. Amano, *Ordered biclique partitions and communication complexity problems*, Discrete Applied Mathematics **184** (2015), 248–252.s

[45] A. Shraibman, *The corruption bound, log-rank, and communication complexity*, Information Processing Letters **141** (2019), 16–21.

[46] M. Sinha and R. de Wolf, *Exponential separation between quantum communication and logarithm of approximate rank*, in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society, Los Alamitos, CA, pp. 966–981.

[47] P. Turán, *On an extremal problem in graph theory*, Matematikai és Fizikai Lapok **48** (1941), 436–452.