Hard Submatrices for Non-Negative Rank and Communication Complexity

Pavel Hrubeš
Institute of Mathematics of ASCR, Prague, Czech Republic

Abstract

Given a non-negative real matrix M of non-negative rank at least r, can we witness this fact by a small submatrix of M? While Moitra (SIAM J. Comput. 2013) proved that this cannot be achieved exactly, we show that such a witnessing is possible approximately: an $m \times n$ matrix of non-negative rank r always contains a submatrix with at most r^3 rows and columns with non-negative rank at least $\Omega(\frac{r}{\log n \log m})$. A similar result is proved for the 1-partition number of a Boolean matrix and, consequently, also for its two-player deterministic communication complexity. Tightness of the latter estimate is closely related to the log-rank conjecture of Lovász and Saks.

2012 ACM Subject Classification Theory of computation

Keywords and phrases Non-negative rank, communication complexity, extension complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2024.13

Funding *Pavel Hrubeš*: This work was supported by Czech Science Foundation GAČR grant 19-27871X.

Acknowledgements The author wants to thank Anup Rao for useful comments.

1 Introduction

The rank of a matrix is one of the most versatile concepts from linear algebra. A basic property of matrix rank is the following: if a matrix M has rank at least r then it contains an $r \times r$ submatrix of rank r. Put differently, the fact that $\mathrm{rk}(M) \geq r$ can be witnessed by a hard $r \times r$ submatrix. Can we extend this witnessing property to other matrix complexity measures? We will consider two such measures: the *non-negative rank* of a non-negative real matrix and the *1-partition number* of a Boolean matrix.

Given a matrix with non-negative real entries, its non-negative rank is defined similarly to rank, except that we want to express the matrix as a sum of non-negative rank-one matrices. This quantity has numerous applications in communication complexity and linear optimization [20], and other fileds (cf. [15]). In [20], Yannakakis has discovered a geometric interpretation of non-negative rank in terms of linear projections of polytopes. This connection has been extended and exploited in many subsequent results, see, e.g., [18, 2, 5], including the current paper.

If M is a $0,1$-matrix, its 1-partition number can be defined as the smallest r such that M can be written as a sum of r rank-one Boolean matrices. This is an important concept in communication complexity [11, 16]. Interpreting a $0,1$-matrix as the adjacency matrix of a bipartite graph, it is also equivalent to the *biclique partition number* (see [3] and references within).

If M has non-negative rank $\geq r$, can this fact be witnessed by a small submatrix? The short answer is no. In [15], Moitra presented an $n \times n$ matrix M of non-negative rank 4 such that every submatrix with less than $n/3$ columns has non-negative rank at most 3 – in particular, M contains no constant-size submatrix of non-negative rank 4. In Section 6.3, we will give a different example where the gap is more dramatic. Similarly, we will see that the most optimistic form of witnessing fails for 1-partition number. On the positive side, we will

39th Computational Complexity Conference (CCC 2024).
Editor: Rahul Santhanam; Article No. 13; pp. 13:1–13:12
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

show that a weaker form of witnessing nevertheless holds: if a matrix has non-negative rank r then it contains a submatrix of size bounded by a polynomial in r whose non-negative rank is close to r; similarly for 1-partition number.

The two-player deterministic communication complexity of M can be characterized by the logarithm of the 1-partition number of M. Hence our witnessing result for 1-partition number can be restated in the language of communication complexity: if a Boolean function has a large communication complexity, this fact can be approximately witnessed by a relatively small set of inputs. It should be noted that this statement immediately follows from the log-rank conjecture of Lovász and Saks (presented in [14]). This conjecture relates the communication complexity of a Boolean matrix with its rank. It implies that for a Boolean matrix M, the three parameters – rank, 1-partition number, non-negative rank – are essentially the same, with their logarithm being polynomially related to the communication complexity of M. This allows us deduce a witnessing property for these measures from the witnessing property of matrix rank. Our result to confirms this prediction of the conjecture and it may therefore be interpreted as a vote in its favor. On the other hand, the log-rank conjecture implies a stronger form of witnessing than what we actually prove. Hence, in principle, a counterexample to the conjecture may be given by a matrix for which this predicted form of witnessing fails (see Section 5 for more details). According to [6], the witnessing problem for communication complexity has been previously posed by H. Halemi.

Our witnessing results could be easily converted to non-trivial *approximation* algorithms to compute non-negative rank or the 1-partition number. These algorithms would run in polynomial time whenever the complexity parameter in question is fixed. Interestingly, *exact* algorithms of this form were given by Moitra [15] and Chandran et al. [3]. While there are similarities between these algorithms and the witnessing perspective, these algorithms ultimately do not search for a witness.

On a more abstract level, the witnessing problem can be posed with respect to any complexity measure whatsoever. A related result in Boolean circuit complexity are "antícheckers" of Lipton and Young [13]. In their work, it is shown that if a Boolean function f requires a Boolean circuit of size s then there is a subset of inputs of size roughly s such that f restricted to this subset still requires circuit size roughly s. A related topic are "hard-core predicates" of Impagliazzo [10]. Recently, Göös et al. [6] studied deterministic query complexity from this perspective. An example from the opposite side of the spectrum is the chromatic number of a graph. It is known that a large chromatic number imposes almost no local structure on a graph and cannot be witnessed by a small subgraph [4, 17].

2 Main results

Given an $m \times n$ matrix M with real non-negative entries, its *non-negative rank*, $\mathrm{rk}_+(M)$, is the smallest s such that M can be written as

$$M = LR\,,$$

where L and R are non-negative matrices of dimensions $m \times s$ and $s \times n$, respectively.

We will show that every M with large non-negative rank contains a relatively small submatrix of large non-negative rank.

▶ **Theorem 1.** *Let M be an $m \times n$ non-negative real matrix with $n \geq 2$. Then for every $k \leq n$, M contains an $m \times k$ submatrix of k columns with non-negative rank $\Omega(R)$, where $R := \min\left((\frac{k}{\log n})^{\frac{1}{3}}, \frac{rk_+(M)}{\log n}\right)$.*

A remarkable consequence is the following:

- M contains an $s_1 \times s_2$ submatrix with $s_1, s_2 \leq \mathrm{rk}_+(M)^3$ and non-negative rank $\Omega(\frac{\mathrm{rk}_+(M)}{\log n \log m})$. Moreover, If M is a square matrix then so is the submatrix.

In some cases, a stronger conclusion is possible. For example, if $\mathrm{rk}_+(M) = n$ then *every* $m \times k$ submatrix of M has non-negative rank k. Theorem 1 becomes interesting if $\log n \ll \mathrm{rk}_+(M) \ll n$. For example, if M is $n \times n$ with $\mathrm{rk}_+(M)$ roughly n^ϵ, we obtain an $n^{3\epsilon} \times n^{3\epsilon}$ submatrix of non-negative rank roughly n^ϵ, and also an $n^\epsilon \times n^\epsilon$ submatrix of non-negative rank roughly $n^{\epsilon/3}$. How far from truth is the estimate from Theorem 1 is an interesting question. In Section 6.3, we will see that the result gives a qualitatively correct picture: the exponent 1/3 can be replaced by 1/2 at best.

Given a Boolean matrix $M \in \{0,1\}^{m\times n}$, let us define its *1-partition number*, $\chi_1(M)$, as the smallest s such that M can be written as

$$M = LR\,, \quad \text{with } L \in \{0,1\}^{m\times s}\,, R \in \{0,1\}^{s\times m}\,,$$

where the operations are over $\mathbb{R}$. The definition emphasizes the analogy with rk_+, and χ_1 is also sometimes referred to as *binary rank*. On the other hand, the phrase "partition number" comes from communication complexity. The name is justified: it is easy to see that $\chi_1(M)$ equals the smallest s such that the 1-entries of M can be partitioned into s 1-monochromatic rectangles (i.e., rank-one Boolean matrices). Finally, when M is viewed as the adjacency matrix of a bipartite graph, $\chi_1(M)$ also appears under the name *biclique partition number* [3].

In the case of χ_1, we obtain a similar but simpler result:

▶ **Theorem 2.** *Let M be an $m \times n$ Boolean matrix with $n \geq 2$. Then for every $k \leq n$, M contains an $m \times k$ submatrix of k columns with 1-partition number $\Omega(\min(\sqrt{k}, \frac{\chi_1(M)}{\log n}))$.*

One consequence is the following (cf. Corollary 6):

- if $\chi_1(M) = p$ then M contains a $p \times p$ submatrix with 1-partition number $\Omega(p^{1/4})$.

The results on 1-partition number imply similar statements in communication complexity; they will be presented in Section 5. Whether these witnessing results can be significantly improved is an intriguing question. It is intimately related to the log-rank conjecture; this connection is discussed in Section 5.

Theorems 1 and 2 are proved in Sections 6.2 and 4, respectively. The proof of Theorem 2 is self-contained. Theorem 1 uses geometrical interpretation of non-negative rank in terms of extended formulations of polytopes and also employs known bounds on complexity of quantifier elimination.

Notation

All logarithms are in base 2 and $[n] := \{1, \ldots, n\}$.

3 A combinatorial lemma

Both Theorems 1 and 2 rely on a simple combinatorial lemma.

▶ **Lemma 3.** *Let $\mathcal{A} \subseteq 2^{[n]}$ be a family of subsets of $[n]$. Assume that $1 \leq k \leq n$ is such that every k-element subset of $[n]$ is contained in some $A \in \mathcal{A}$. Then there exists a subfamily $\mathcal{A}' \subseteq \mathcal{A}$ of size $|\mathcal{A}'| \leq O(|\mathcal{A}|^{\frac{1}{k}} \log(n/k))$ with $\bigcup \mathcal{A}' = [n]$. In particular, if $|\mathcal{A}| \leq 2^k$ then $|\mathcal{A}'| \leq O(\log n)$.*

Proof. Assume that $|\mathcal{A}| \leq a^k$. Let t be the size of a largest set in $\mathcal{A}$. Then we have

$$\binom{n}{k} \leq a^k \binom{t}{k} .$$

Hence $t \geq \frac{n}{ea}$, using the estimates $(\frac{n}{k})^k \leq \binom{n}{k}$, $\binom{t}{k} \leq (\frac{et}{k})^k$. Take some $A_0 \in \mathcal{A}$ of size t. Let

$$\mathcal{A}_1 := \{A \setminus A_0 : A \in \mathcal{A}\} .$$

Then every subset of $U_1 := [n] \setminus A_0$ of size *at most* k is contained in some member of $\mathcal{A}_1$. The size of U_1 is at most $n(1 - \frac{1}{ea})$. Similarly, take a largest set A_1 from $\mathcal{A}_1$ and obtain a new family $\mathcal{A}_2 \subseteq 2^{U_2}$ on $U_2 := U_1 \setminus A_1$. After s steps, the size of U_s is at most $n(1 - \frac{1}{ea})^s$ and after $s \leq O(a \log(n/k))$ steps we have $|U_s| \leq k$. This guarantees that the largest set in $\mathcal{A}_s$ is U_s itself and $[n] = \bigcup_{i=0}^{s} A_s$. By construction, every A_i is contained in some element of the original family $\mathcal{A}$. ◀

For some range of parameters, the lemma can be also proved from the Min Max Theorem of Lipton and Young in [13] which would also give an approximate version of it.

An application (which will not be explicitly used) is the following. A subadditive measure on $[n]$ is a function $\mu : 2^{[n]} \to \mathbb{R}$ such that $\mu(A_1 \cup A_2) \leq \mu(A_1) + \mu(A_2)$ holds for every $A_1, A_2 \subseteq [n]$.

▶ **Corollary 4.** *Let μ be a subadditive measure on $[n]$. Assume $1 \leq k \leq n$ and that every k-element subset of $[n]$ has measure at most s. Let N be the number of $\subseteq$-maximal subsets of $[n]$ of measure at most s. Then $\mu([n]) \leq O(sN^{\frac{1}{k}} \log(n/k)))$.*

4 1-Partition number

In this section, we prove Theorem 2.

Let M be an $m \times n$ matrix with rows indexed by $[n] = \{1, \ldots, n\}$. Given $A \subseteq [n]$, M_A denotes the submatrix obtained by removing the rows outside of A from M. Observe that[1]

$$\chi_1(M_{A_1 \cup A_2}) \leq \chi_1(M_{A_1}) + \chi_1(M_{A_2}) , \tag{1}$$

and so $\chi_1(M_A)$ can be viewed as a subadditive measure on $[n]$ whenever M is fixed.

If a matrix M has rank r, its rows are a linear combination of a subset of r rows of M. This means that every column of M is determined by a fixed subset of r coordinates. If M is Boolean, this leads to the following useful fact:

- if M has distinct columns then $n \leq 2^{\mathrm{rk}(M)}$ (similarly for rows).

▶ **Lemma 5.** *Let M be an $m \times n$ Boolean matrix of rank r. Given $s \in [n]$, let $\mathcal{A}$ be the collection of maximal subsets $A \subseteq [n]$ with $\chi_1(M_A) \leq s$ (i.e., $\chi_1(M_A) \leq s$ and $\chi_1(M_{A'}) > s$ for every $A' \supsetneq A$). Then $|\mathcal{A}| \leq 2^{(r+s)^2}$.*

Proof. Let $v_1, \ldots, v_n \in \mathbb{R}^m$ be the columns of M. Given $L \in \{0,1\}^{m \times s}$, let

$$L^* := \{i \in [n] : \exists y \in \{0,1\}^s \, v_i = Ly\} .$$

Let $\mathcal{L} := \{L^* : L \in \{0,1\}^{m \times s}\}$.

[1] If A_1, A_2 are disjoint, this is quite obvious. Otherwise consider A_1, $A_2 \setminus A_1$.

We claim that $\mathcal{A} \subseteq \mathcal{L}$. If $\chi_1(M_A) \leq s$, we can write $M_A = LR$ with $L \in \{0,1\}^{m\times s}$ and $R \in \{0,1\}^{s\times|A|}$. This means that every v_i, $i \in A$, is a Boolean linear combination of the columns of L and $A \subseteq L^*$. Furthermore, if A is maximal, we must have $A = L^*$.

We now want to estimate the size of $\mathcal{L}$. The set L^* consists of indices $i \in [n]$ so that there exists $x \in \mathbb{R}^n$, $y \in \mathbb{R}^s$ satisfying

$$Mx - Ly = 0 \tag{2}$$

such that $y \in \{0,1\}^s$ and x is the i-th unit vector. Since M has rank r and L has rank at most s, the system (2) is equivalent to a subsystem of $t := \min((s+r), m)$ equations. These correspond to rows of the matrix (M, L). Hence, in order to determine L^*, it is sufficient to specify a t-element subset of $[m]$ together with the $t \times s$ submatrix of L. This gives the estimate

$$|\mathcal{L}| \leq \binom{m}{t} 2^{ts} \leq 2^{t(s+\log m)} .$$

Finally, we can assume that M has distinct rows and so $\log m \leq r$, obtaining the bound $2^{(r+s)^2}$. ◀

▶ **Theorem 2** (restated). *Let M be an $m\times n$ Boolean matrix with $n \geq 2$. Then for every $k \leq n$, M contains an $m \times k$ submatrix of k columns with 1-partition number $\Omega(\min(\sqrt{k}, \frac{\chi_1(M)}{\log n}))$.*

Proof. Let r be the rank of M. We will assume $r \leq \frac{k^{1/2}}{2}$. Otherwise, observe that M contains a full rank $r \times r$ submatrix, χ_1 is lower-bounded by rank, and the conclusion of the theorem follows.

Let s be the maximum $\chi_1(M_A)$ over all $A \subseteq [n]$ of size k. Let $\mathcal{A}$ be the family from the previous lemma. If $|\mathcal{A}| \geq 2^k$, we have $2^k \leq 2^{(s+r)^2}$ and therefore $s \geq \frac{k^{1/2}}{2}$ from the assumption on r.

Assume $|\mathcal{A}| \leq 2^k$. By Lemma 3, there exists a subfamily $\mathcal{A}' \subseteq \mathcal{A}$ of size $O(\log n)$ which covers $[n]$. Using (1), this implies $\chi_1(M) \leq O(s \log n)$ and so $s \geq \Omega(\chi_1(M)/\log n)$. ◀

▶ **Corollary 6.** *Let M be as above with $\chi_1(M) = p$. Then M contains*
- **(i)** *a submatrix of at most p^2 columns with partition number $\Omega(p/\log n)$,*
- **(ii)** *a submatrix with at most p^2 rows and columns with partition number $\Omega(p/(\log n \log m))$. If M is a square matrix then so is the submatrix.*
- **(iii)** *a submatrix with p columns with partition number $\Omega(p^{\frac{1}{2}})$*
- **(iv)** *a $p \times p$ submatrix with partition number $\Omega(p^{\frac{1}{4}})$.*

Proof. Part (i). If $n \leq p^2$, M itself satisfies the statement. Otherwise apply the theorem with $k = p^2$.

Part (ii). Apply (i) again to the transpose of the submatrix obtained in (i). If $m = n$, we can enlarge the submatrix to a square matrix.

Part (iii). Without loss of generality, we can assume that the columns of M are distinct. This implies that M has rank at least $\log n$. If $\sqrt{p} \leq p/\log n$, apply the theorem to obtain the desired matrix. Otherwise, we have $p \geq \log^2 n$. M contains a submatrix of p columns of rank at least $\min(p, \log n) \geq \sqrt{p}$.

Part (iv) follows by taking the submatrix from (iii), and applying (iii) to its transpose. ◀

▶ Remark 7. The bound of Theorem 2 can be slightly improved to give $\Omega(\sqrt{k \log(1 + \frac{\chi_1(M)}{k^{1/2}\log n})})$, as long as $k^{1/2} \leq \chi_1(M)/\log n$. For example, if $k = \chi_1(M)/\log n$, we obtain a submatrix of k columns with 1-partition number $\Omega(\sqrt{k \log k})$.

Furthermore, M always contains a submatrix M' of k columns with $\chi_1(M') \geq \chi_1(M) \cdot \lceil \frac{n}{k} \rceil^{-1}$, which gives better parameters if $\chi_1(M)$ is close to n.

4.1 A somewhat non-trivial example

We now give a finite example which shows that the most optimistic form of witnessing fails for χ_1.

▶ **Theorem 8.** *There exists a* 5×6 *Boolean matrix* M *with* $\chi_1(M) = 5$ *such that every* 5×5 *submatrix of* M *has 1-partition number at most* 4.

Proof. Let

$$M := \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} .$$

We first argue that $\chi_1(M) > 4$, which implies $\chi_1(M) = 5$ since M has 5 rows.

Suppose that $\chi_1(M) \leq 4$. Then there exists a set of Boolean row-vectors $V = \{v_1, \ldots, v_4\}$ such that every row of M is their Boolean linear combination; i.e., of the form $\sum_{i \in A} v_i$ for some $A \subseteq \{1, \ldots, 4\}$. Note that in this expression, the non-zero coordinates of v_i, $i \in A$, are a subset of the non-zero coordinates of the given row. Using this observation, it is easy to see that V must consist of the first 4 rows of M. If $\chi_1(M) \leq 4$ this means that the last row of M is a Boolean combination of the first four rows, which is clearly impossible.

We now show that every submatrix obtained by removing a column from M has χ_1 at most 4.

First, assume that M' has been obtained by removing the third column. The resulting matrix, together with a partition into four 1-monochromatic rectangles a, b, c, d, is as follows:

$$M' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} , \quad \begin{pmatrix} a & 0 & 0 & a & b \\ 0 & c & c & 0 & b \\ 0 & 0 & d & d & 0 \\ 0 & 0 & d & d & b \\ a & c & c & a & b \end{pmatrix} .$$

Second, assume that M'' has been obtained by removing the last column. The resulting matrix, together with its partition, is the following:

$$M'' = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} , \quad \begin{pmatrix} a & 0 & 0 & 0 & a \\ 0 & b & 0 & b & 0 \\ 0 & 0 & c & d & d \\ 0 & 0 & 0 & d & d \\ a & b & c & b & a \end{pmatrix} .$$

Finally, note that if we remove from M the first or the second column, we obtain M' (up to a permutation of rows and columns). And, if we remove the fourth or fifth column, we obtain M''. Hence indeed, every 5×5 submatrix has χ_1 at most 4 ◀

By placing n copies of the matrix from Theorem 8 on diagonal, we obtain:

▶ **Corollary 9.** *For every* n*, there exists a* $5n \times 6n$ *Boolean matrix* M *with* $\chi_1(M) = 5n$ *such that every submatrix obtained by removing a column of* M *has* 1*-partition number strictly less than* $5n$.

5 Communication complexity, and a comparison with the log-rank conjecture

Given an $m \times n$ Boolean matrix M, consider the following two-player game: Alice knows $i \in [m]$, Bob knows $j \in [n]$, and they are supposed to compute the value of $M_{i,j}$. Denote by $\mathrm{cc}(M)$ the deterministic communication complexity of this game. For details about the communication model, see for example [11, 16].

In order to relate communication complexity with χ_1, we need the following classical fact (the first inequality is due to Yao, the second is due to Yannakakis [20]): if M is non-constant then

$$\log(\chi_1(M) + 1) \le \mathrm{cc}(M) \le O(\log^2 \chi_1(M))\,. \tag{3}$$

▶ **Proposition 10.** *Let M be a Boolean matrix with communication complexity c. Then there exist $k \ge \Omega(\sqrt{c})$ and a $2^k \times 2^k$ submatrix of M with communication complexity at least $k/4 - O(1)$.*

Proof. From (3), there exists $k \ge \Omega(\sqrt{c})$ with $\chi_1(M) \ge 2^k$. Corollary 6, part (iv), gives $2^k \times 2^k$ submatrix M' with $\chi_1(M') \ge \Omega(2^{k/4})$. By (3), we have $\mathrm{cc}(M') \ge k/4 - O(1)$. ◀

It is worthwhile to compare this with what is predicted by the log-rank conjecture [14] of Lovász and Saks.

▶ **Log-rank conjecture.** *There is a constant α such that $cc(M) \le O(\log^\alpha(rk(M)))$ for any non-zero Boolean matrix M.*

▶ **Proposition 11.** *Assume the log-rank conjecture. Then every Boolean matrix with communication complexity c contains a $2^k \times 2^k$ submatrix M' with $\chi_1(M') = 2^k$, communication complexity $k+1$, and $k \ge \Omega(c^{1/\alpha})$.*

Proof. If M has communication complexity c then, by the log-rank conjecture, M has rank at least 2^k with $k \ge \Omega(c^{1/\alpha})$. Hence M contains a full-rank $2^k \times 2^k$ submatrix M'. Since $\chi_1(M') \ge \mathrm{rk}(M')$, we have $\chi_1(M') = 2^k$. If c is sufficiently large, so that $k \ge 1$, then M' is non-constant and we obtain $\mathrm{cc}(M') \ge k+1$ by (3). ◀

This is almost what has been proved in Proposition 10. One difference is that the constant α in Proposition 11 is unconditionally set to 2 in Proposition 10. However, there is a more important qualitative difference. The submatrix presented in Proposition 11 has *highest possible* communication complexity: the protocol in which Alice sends her input to Bob and Bob sends back the answer (or vice versa), is optimal. Any other protocol cannot save even *one bit* of communication. In contrast, Proposition 10 presents a submatrix with only a very high communication complexity. To summarize, Proposition 10 confirms a prediction of the log-rank conjecture. But with worse parameters than what the conjecture predicts: consequently the bound in the proposition is far from from tight, or the conjecture is false.

Another consequence is:

▶ Remark 12. In order to solve the log-rank conjecture, it is sufficient to focus on $2^k \times 2^k$ matrices with communication complexity at least $k/4 - O(1)$.

6 Non-negative rank

6.1 Extended formulations and separation complexity

Let us first make a short detour into extended formulations of convex polyhedra.

A *polyhedron* $P \subseteq \mathbb{R}^r$ is a (possibly unbounded) set defined by a finite number of linear constraints. Following [20, 18, 2], define the *extension complexity of* P, $\mathrm{xc}(P)$, as the smallest s such that P is a linear projection of a polyhedron $Q \subseteq \mathbb{R}^m$ where Q can be defined using s inequalities (and any number of equalities). Observe that P with extension complexity s can be expressed in the standard form

$$x \in P \text{ iff } \exists_{y \in \mathbb{R}^s}\, Cx + Dy = b, y \geq 0\,,$$

where $C \in \mathbb{R}^{t \times r}, D \in \mathbb{R}^{t \times s}$ and $b \in \mathbb{R}^t$ for some t.

Let V be a finite subset of $\mathbb{R}^r$. Given $A \subseteq V$, its *separation complexity*, $\mathrm{sep}_V(A)$, is the minimum $\mathrm{xc}(P)$ over all polyhedra $P \subseteq \mathbb{R}^r$ with[2]

$$P \cap V = A\,;$$

such a P is called a *separating* polyhedron for A. In other words, $\mathrm{sep}_V(A)$ is the smallest s so that we can distinguish points in A from points in $V \setminus A$ by means of a linear program with s inequalities. Moreover, such a program can be rewritten as

$$x \in A \text{ iff } (\, x \in V \text{ and } \exists_{y \in \mathbb{R}^s}\, Cx + Dy = b, y \geq 0\,)\,.$$

The notion of separation complexity has been studied in [7, 8, 9] in the case when $V = \{0,1\}^n$ is the Boolean cube. The following theorem is of independent interest and can be seen as an extension of similar results in [7, 9]. The proof is a considerable simplification of the previous ones.

▶ **Theorem 13.** *Let V be a non-empty finite subset of $\mathbb{R}^r$. Given a parameter $s \geq 1$, let $\mathcal{A}$ be the collection of subsets A of V with $\mathrm{sep}_V(A) \leq s$. Then*

$$|\mathcal{A}| \leq 2^{O(s(r+s)^2 \log |V|)}\,.$$

The proof is delegated to the appendix.

An immediate consequence of Theorem 13 is a theorem from [9]:

- if $V = \{0,1\}^n$ then there exists $A \subseteq V$ with $\mathrm{sep}_V(A) \geq 2^{n^{\frac{1}{3}(1-o(1))}}$.

6.2 Submatrices of large non-negative rank

In order to apply Theorem 13, we also need a connection between extension complexity and non-negative rank. This is provided by the notion of slack matrix introduced in [20]. Following [20, 2], we now define what it is. Let V be a sequence $v_1, \ldots, v_{m_1}$ of points in $\mathbb{R}^r$ and $L(x)$ a system $\ell_1(x) \geq b_1, \ldots, \ell_{m_2}(x) \geq b_{m_2}$ of inequalities in $\mathbb{R}^r$. The slack matrix with respect to V and $L(x)$ is the $m_2 \times m_1$ matrix S such that

$$S_{i,j} = \ell_i(v_j) - b_i\,.$$

Let $P_0 := \mathrm{conv}(V)$ be the convex hull of V and $P_1 := \{x \in \mathbb{R}^r :\ L(x) \text{ holds}\}$. If $P_0 \subseteq P_1$ then S is non-negative. In [2], we can find:

[2] If no such polyhedron exists, which may happen if V is not convexly independent, we set $\mathrm{sep}_V(A) := \infty$.

▶ **Lemma 14** ([2])**.** *Let $P_0 \subseteq P_1$ and S be as above. Define $xc(P_0, P_1)$ as the minimum $xc(P)$ over all polyhedra with $P_0 \subseteq P \subseteq P_1$. Then*

$$rk_+ S - 1 \le xc(P_0, P_1) \le rk_+ S \,.$$

▶ **Theorem 1** (restated)**.** *Let M be an $m \times n$ non-negative real matrix with $n \ge 2$. Then for every $k \le n$, M contains an $m \times k$ submatrix of k columns with non-negative rank $\Omega(R)$, where $R := \min\left((\frac{k}{\log n})^{\frac{1}{3}}, \frac{rk_+(M)}{\log n}\right)$.*

Proof. Let r be the rank of M. We can write $M = LR$ where $L \in \mathbb{R}^{m\times r}$, $R \in \mathbb{R}^{r\times n}$. Let $V \subseteq \mathbb{R}^r$ be the set of columns $v_1, \dots v_n$ of R. (Without loss of generality, the columns of M are distinct). Given $A \subseteq [n]$, let M_A be the submatrix obtained by deleting columns outside of A from M. Also let $V_A := \{v_i : i \in A\}$. Then M_A can be interpreted as the slack matrix of the polytope $P_A = \mathrm{conv}(V_A)$ and the polyhedron $Q = \{x \in \mathbb{R}^d : Lx \ge 0\}$.

Suppose that for every A of size k, $\mathrm{rk}_+(M_A) \le s$. Then for every such A, there is a polyhedron Q_A with $V_A \subseteq Q_A \subseteq Q$ with $\mathrm{xc}(Q_A) \le s$. Let $A^* := V \cap Q_A$. Then Q_A is a separating polyhedron for $A^* \supseteq A$. Let $\mathcal{A}$ be the collection of A^* over all A of size k. Theorem 13 implies

$$|\mathcal{A}| \le 2^{c \log n (s+r)^3} \,,$$

where c is an absolute constant.

We will assume $r \le (\frac{k}{2c\log n})^{1/3}$. Otherwise M contains a full rank $r \times r$ submatrix, rk_+ is lower-bounded by rank, and the conclusion of the theorem follows.

If $|\mathcal{A}| \ge 2^k$, we obtain $c \log n(s+r)^3 \ge k$ and hence $s \ge \Omega((k/\log n)^{1/3})$ from the assumption on r.

Assume $|\mathcal{A}| \le 2^k$. By Lemma 3, there exists a subfamily $\mathcal{A}' \subseteq \mathcal{A}$ of size $O(\log n)$ which covers $[n]$. This implies (note that (1) holds also for non-negative rank) $\mathrm{rk}_+(M) \le O(s \log n)$ and $s \ge \Omega(\mathrm{rk}_+(M)/\log n)$. ◀

The following is proved similarly to Corollary 6:

▶ **Corollary 15.** *Let M be a non-negative $m \times n$ matrix with $rk_+(M) = p$. Then M contains*

(i) *an $s_1 \times s_2$ submatrix with $s_1, s_2 \le p^3$ with non-negative rank $\Omega(\frac{p}{\log n \log m})$. If $m = n$, we can assume $s_1 = s_2$.*

(ii) *a $p \times p$ submatrix with non-negative rank $\Omega(\frac{p^{\frac{1}{3}}}{\log^{\frac{1}{3}} n \log m})$.*

6.3 Tightness

In [15], Moitra has constructed a non-negative matrix M with the following properties:

- M is $3rn \times 3rn$, $\mathrm{rk}_+(M) \ge 4r$, any submatrix with $< n$ columns has non-negative rank at most $3r$.

Observe that in order to witness the non-negative rank of this M exactly, one needs a constant fraction of the columns of M. On the other hand, the gap between the non-negative rank of M and that of its submatrices is quite mild.

We now give a different example which is of a similar flavor as the bound from Theorem 1. It also shows that the constant $\frac{1}{3}$ in the theorem can be replaced by $\frac{1}{2}$ at best. The example follows from very non-trivial results of Kwan et al. [12]. A similar bound would follow from the more general result of Shitov [19].

▶ **Theorem 16.** *For every n, there exists an $n \times n$ matrix with non-negative rank $\Omega(\sqrt{n})$ such that every $n \times k$ submatrix has non-negative rank $O(\sqrt{k})$.*

Proof. From [12], there exists an n-vertex polygon $P \subseteq \mathbb{R}^2$ with vertices lying on the unit circle with extension complexity $\Omega(\sqrt{n})$. Let M be its slack matrix with columns corresponding to vertices $v_1, \dots, v_n$ of P. From Lemma 14, we have $\mathrm{rk}_+(M) \geq \Omega(\sqrt{n})$. Given an $n \times k$ submatrix M' with columns $i_1, \dots, i_k$, Lemma 14 shows that $\mathrm{rk}_+(M')$ is at most the extension complexity of $\mathsf{conv}(v_{i_1}, \dots, v_{i_k})$ (plus 1). Using another result from [12], every k-gon with vertices on the unit circle has extension complexity at most $O(\sqrt{k})$. ◀

7 Open problems

Our first two open problems are concerned with tightness of the bounds in Theorems 1 and 2.

▶ **Open problem 1.** *Let M be $m \times n$ non-negative matrix. Does M contain a submatrix of at most $rk_+(M)^2$ columns with non-negative rank $\Omega(rk_+(M))$?*

▶ **Open problem 2.** *Find a Boolean matrix M with $\chi_1(M) = p$ such that every $p \times p$ submatrix has 1-partitition number much smaller than p.*

As far as we can see, the bound from Problem 1 is consistent with what we know about non-negative rank, and would be optimal. The task is to improve Corollary 15(i) in two different ways: first, to reduce the dependence on $\mathrm{rk}_+(M)$ from cubic to quadratic and, second, to eliminate the logarithmic dependence on the size of M altogether. For Problem 2, Theorem 8 gives an M with submatrices of χ_1 strictly less than p; there should exist a construction with a larger gap.

As discussed in Section 5, in order to solve the log-rank conjecture, it is enough to focus on matrices with large 1-partition number. The following is the extreme case of this question:

▶ **Open problem 3.** *Suppose M is $n \times n$ Boolean matrix with $\chi_1(M) = n$. How small can the rank of M be?*

References

1. S. Basu, R. Pollack, and M.F. Roy. *Algorithms in real algebraic geometry*. Springer-Verlag, 2006.
2. G. Braun, S. Fiorini, S. Pokutta, and D. Steuer. Approximation limits of linear programs (beyond hierarchies). *Math. of Operations Research*, 40(3), 2015.
3. S. Chandran, D. Issac, and A. Karrenbauer. On the parameterized complexity of biclique cover and partition. In *International Symposium on Parameterized and Exact Computation*, 2016.
4. P. Erdös. Problems and results in combinatorial analysis and graph theory. *Annals of Discrete Mathematics*, 38:81–92, 1988.
5. S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Linear vs. semidefinite extended formulations: exponential separation and strong lower bounds. In *Symposium on the Theory of Computing*, 2011.
6. M. Göös, I. Newman, A. Riazanov, and D. Sokolov. Hardness condensation by restriction. *ECCC*, 2023.
7. P. Hrubeš. On ϵ-sensitive monotone computations. *Computational Complexity*, 29(2), 2020.
8. P. Hrubeš. On the complexity of computing a random Boolean function over the reals. *Theory of Computing*, 16(9):1–12, 2020.
9. P. Hrubeš and N. Talebanfard. On the extension complexity of polytopes separating subsets of the Boolean cube. *Disc. and Comp. Geom.*, 70(1), 2022.

10 R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 538–545, 1995.
11 Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996.
12 Matthew Kwan, Lisa Sauermann, and Yufei Zhao. Extension complexity of low-dimensional polytopes. *Transactions of the American Mathematical Society*, 375(6), 2022.
13 Richard Lipton and Neal Young. Simple strategies for large zero-sum games with applications to complexity theory. *CoRR*, cs.CC/0205035, May 2002. doi:10.1145/195058.195447.
14 L. Lovász and M. Saks. Lattices, mobius functions and communications complexity. In *29th Annual Symposium on Foundations of Computer Science (FOCS 1988)*, pages 81–90, 1988.
15 Ankur Moitra. An almost optimal algorithm for computing nonnegative rank. In *SIAM J. Comput.*, 2013. URL: https://api.semanticscholar.org/CorpusID:1920044.
16 Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.
17 V. Rödl and R. A. Duke. On graphs with small subgraphs of large chromatic number. *Graphs and Combinatorics*, 1:91–96, 1985.
18 Thomas Rothvoß. Some 0/1 polytopes need exponential size extended formulations. *CoRR*, abs/1105.0036, 2011. arXiv:1105.0036.
19 Y. Shitov. Sublinear extension of polygons. *arXiv*, 2014. arXiv:1412.0728.
20 Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991.

A Proof of Theorem 13

The proof uses known results on quantifier elimination which we first outline. We follow the monograph of Basu, Pollack and Roy [1]. Theorem 13 requires an elimination of only a single block of existential quantifiers, so we focus on this case only.

For $b \in \mathbb{R}$, let

$$\mathrm{sgn}(b) := \begin{cases} 1\,, & b > 0\,,\\ 0\,, & b = 0\,,\\ -1\,, & b < 0\,. \end{cases}$$

Given $b = \langle b_1, \dots, b_m \rangle \in \mathbb{R}^m$, let $\mathrm{sgn}(b) := \langle \mathrm{sgn}(b_1), \dots, \mathrm{sgn}(b_m) \rangle \in \{-1, 0, 1\}^m$. Let $F = F(z, y)$ be a sequence of m polynomials $f_1, \dots, f_m \in \mathbb{R}[z, y]$ in variables $z = \{z_1, \dots, z_{k_1}\}$ and $y = \{y_1, \dots, y_{k_2}\}$. Given $a \in \mathbb{R}^{k_1}$, define $\mathrm{SGN}_1(F, a) \subseteq \{-1, 0, 1\}^m$

$$\mathrm{SGN}_1(F, a) := \{\mathrm{sgn}(F(a, b)) : b \in \mathbb{R}^{k_2}\}\,.$$

Let

$$\mathrm{SGN}(F) := \{\mathrm{SGN}_1(F, a) : a \in \mathbb{R}^{k_1}\}\,.$$

Theorem 14.16 from [1] provides the following bound on the size of SGN:

▶ **Theorem** ([1]). *If every polynomial in F has degree at most d then*

$$|\mathrm{SGN}(F)| \le m^{(k_1+1)(k_2+1)} d^{O(k_1)O(k_2)}\,. \tag{4}$$

We now apply this result to the case of Theorem 13. Let $V, s, \mathcal{A}$ be as in the assumption. Every $A \in \mathcal{A}$ can be described by a linear system with s inequalities. Namely, for every $x \in V$,

$$x \in A \text{ iff } \exists_{y \in \mathbb{R}^s}\, Cx + Dy = b,\ y \ge 0\,, \tag{5}$$

where $C \in \mathbb{R}^{t \times r}, D \in \mathbb{R}^{t \times s}$ and $b \in \mathbb{R}^t$. Since $Cx + Dy = b$ is a system of equations in $r + s$ variables x, y, we can also assume $t = r + s$.

Let us view the parameters C, D, b in (5) as variables. Let z be the set of these variables, of size $k_1 = (r+s)(r+s+1)$. Given $v \in V$, let $F_v(z,y)$ be the sequence of $(r+s)$ polynomials

$$Cv + Dy - b$$

in variables z and $y = \{y_1, \ldots, y_s\}$. Let $F(z,y)$ be the union of $F_v(z,y)$ over all $v \in V$, together with the polynomials $y_1, \ldots, y_s$. Hence F consists of $m = s + |V|(r+s)$ polynomials of degree at most two.

$F(z,y)$ is set up so that

$$|\mathcal{A}| \leq |\mathrm{SGN}(F)| \,.$$

To see this, observe that whenever the parameters z are fixed, the set $A \subseteq V$ given by (5) is uniquely determined by $\mathrm{SGN}_1(F(z,y))$. Since every $A \in \mathcal{A}$ is obtained by some fixing of the parameters, we indeed obtain $|\mathcal{A}| \leq |\mathrm{SGN}(F(z,y))|$.

Finally, we can apply (4) to estimate $|\mathrm{SGN}(F)|$ with $m = s + |V|(r+s)$, $k_1 = (r+s)(r+s+1)$, $k_2 = s$, and $d = 2$. To simplify the expression, we can assume $s + r \leq |V|$; otherwise the upper bound asserted in Theorem 13 exceeds the trivial bound $|\mathcal{A}| \leq 2^{|V|}$. This means that $m \leq 2|V|^2$. If we loosen the bound (4) as $|\mathrm{SGN}(F)| \leq (dm)^{O(k_1)O(k_2)}$, we obtain (recall that $s \geq 1$)

$$|\mathrm{SGN}(F)| \leq 2^{O(s(s+r)^2 \log |V|)} \,,$$

as required.