Tight bounds on the maximum size of a set of permutations with bounded VC-dimension *

Josef Cibulka¹, Jan Kynčl²

¹ Department of Applied Mathematics, Charles University, Faculty of Mathematics and Physics, Malostranské nám. 25, 118 00 Praha 1, Czech Republic; cibulka@kam.mff.cuni.cz
² Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Faculty of Mathematics and Physics, Malostranské nám. 25, 118 00 Praha 1, Czech Republic; kyncl@kam.mff.cuni.cz

Abstract

The VC-dimension of a family 𝒫 of 𝑛-permutations is the largest integer 𝑘 such that the set of restrictions of the permutations in 𝒫 on some 𝑘-tuple of positions is the set of all 𝑘! permutation patterns. Let 𝑟_𝑘(𝑛) be the maximum size of a set of 𝑛-permutations with VC-dimension 𝑘. Raz showed that 𝑟_2(𝑛) grows exponentially in 𝑛. We show that 𝑟_3(𝑛) = 2^{Θ(𝑛α(𝑛))} and for every 𝑡 ≥ 1, we have 𝑟_{2t+2}(𝑛) = 2^{Θ(𝑛α(𝑛)^t)} and 𝑟_{2t+3}(𝑛) = 2^{O(𝑛α(𝑛)^t log α(𝑛))}.

We also study the maximum number 𝑝_𝑘(𝑛) of 1-entries in an 𝑛×𝑛 (0,1)-matrix with no (𝑘+1)-tuple of columns containing all (𝑘+1)-permutation matrices. We determine that 𝑝_3(𝑛) = Θ(𝑛α(𝑛)) and 𝑝_{2t+2}(𝑛) = n^{2(1/2)α(𝑛)^t ± O(α(𝑛)^t−1)} for every 𝑡 ≥ 1.

We also show that for every positive 𝑠 there is a slowly growing function ζ_𝑠(𝑚) (for example ζ_5(𝑚) = 2^{O(α(𝑠−3)/2(𝑚))}) for every odd 𝑠 ≥ 5) satisfying the following. For all positive integers 𝑚, 𝑛, 𝐵 and every 𝑚×𝑛 (0,1)-matrix 𝑀 with ζ_𝑠(𝑚)𝐵𝑛 1-entries, the rows of 𝑀 can be partitioned into 𝑠 intervals so that some ⌊𝐵𝑛/𝑚⌋-tuple of columns contains at least 𝐵 1-entries in each of the intervals.

1 Introduction

Let 𝒯 be a set system on [𝑛] = {1, 2, ..., 𝑛}. We say that a set 𝐾 ⊂ [𝑛] is shattered by 𝒯 if every subset of 𝐾 appears as an intersection of 𝐾 and some set from 𝒯. The Vapnik-Chervonenkis dimension (VC-dimension) of 𝒯 is the size of the largest set shattered by

*Research was supported by the project 1M0545 of the Ministry of Education of the Czech Republic, by project no. 52410 of the Grant Agency of Charles University and by the grant SVV-2010-261313 (Discrete Methods and Algorithms). Josef Cibulka was also supported by the Czech Science Foundation under the contract no. 201/09/H057.
Sauer’s lemma gives a tight upper bound on the size of a set system on \([n]\) with VC-dimension \(k\), which is a polynomial in \(n\) of degree \(k\). More on the VC-dimension and its history can be found in [16].

Motivated by the so-called acyclic linear orders problem, Raz [22] defined the VC-dimension of a set \(P\) of permutations: Let \(S_n\) be the set of all \(n\)-permutations, that is, permutations of \([n]\). The restriction of \(\pi \in S_n\) to the \(k\)-tuple \((a_1, a_2, \ldots, a_k)\) of positions (where \(1 \leq a_1 < a_2 < \cdots < a_k \leq n\)) is the \(k\)-permutation \(\pi'\) satisfying \(\forall i, j : \pi'(i) < \pi'(j) \iff \pi(a_i) < \pi(a_j)\). The \(k\)-tuple of positions \((a_1, \ldots, a_k)\) is shattered by \(P\) if each \(k\)-permutation appears as a restriction of some \(\pi \in P\) to \((a_1, \ldots, a_k)\). The VC-dimension of \(P\) is the size of the largest set of positions shattered by \(P\).

Let \(r_k(n)\) be the size of the largest set of \(n\)-permutations with VC-dimension \(k\).

Raz [22] proved that \(r_2(n) \leq C^n\) for some constant \(C\) and asked whether an exponential upper bound on \(r_k(n)\) can also be found for every \(k \geq 3\).

An \(n\)-permutation \(\pi\) avoids a \(k\)-permutation \(\rho\) if none of the restrictions of \(\pi\) to a \(k\)-tuple of positions is \(\rho\). Clearly, the set of permutations avoiding \(\rho \in S_k\) has VC-dimension smaller than \(k\). Thus, Raz’s question generalizes the Stanley–Wilf conjecture which states that the number of \(n\)-permutations that avoid an arbitrary fixed permutation \(\rho\) grows exponentially in \(n\). The conjecture was settled by Marcus and Tardos [15] using a result of Klazar [12].

Let \(\alpha(n)\) be the inverse of the Ackermann function; see Section 2.2 for its definition.

We show in Section 2 that the size of a set of \(n\)-permutations with VC-dimension \(k\) cannot be much larger than exponential in \(n\).

**Theorem 1.**

\[
\begin{align*}
r_3(n) &\leq \alpha(n)^{(4+\alpha(1))n}, \\
r_4(n) &\leq 2^{n(2\alpha(n)+3\log_2(\alpha(n))+O(1))}, \\
r_{2t+2}(n) &\leq 2^{n((2/t)\alpha(n)^t+O(\alpha(n)^{t-1}))} \text{ for every } t \geq 2 \\
r_{2t+3}(n) &\leq 2^{n((2/t)\alpha(n)^t\log_2(\alpha(n))+O(\alpha(n)^t))} \text{ for every } t \geq 1.
\end{align*}
\]

The result has an application in enumerating simple complete topological graphs [14].

On the other hand, we give a negative answer to the Raz’s question:

**Theorem 2.**

\[
r_3(n) \geq (\alpha(n)/2 - O(1))^n.
\]

Let \(k \geq 4\) be a fixed integer and let \(t := [(k-2)/2]\). We have

\[
r_k(n) \geq 2^{n((1/t)\alpha(n)^t-O(\alpha(n)^{t-1}))}.
\]

**Theorem 2** is proved in Section 3.

An \(n\)-permutation matrix is an \(n \times n\) \((0, 1)\)-matrix with exactly one 1-entry in every row and every column. Permutations and permutation matrices are in a one-to-one correspondence that assigns to a permutation \(\pi\) a permutation matrix \(A_\pi\) with \(A_\pi(i, j) = 1 \iff \pi(j) = i\).

An \(m \times n\) \((0, 1)\)-matrix \(B\) contains a \(k \times l\) \((0, 1)\)-matrix \(S\) if \(B\) has a \(k \times l\) submatrix \(T\) that can be obtained from \(S\) by changing some (possibly none) 0-entries to 1-entries.
Otherwise $B$ avoids $S$. Thus, a permutation $\pi$ avoids $\rho$ if and only if $A_{\pi}$ avoids $A_{\rho}$. Füredi and Hajnal [7] studied the following problems from the extremal theory of $(0, 1)$-matrices. Given a matrix $S$ (the forbidden matrix), what is the maximum number $\text{ex}_S(n)$ of 1-entries in an $n \times n$ matrix that avoids $S$? This area is closely related to the Turán problems on graphs and to the Davenport–Schinzel sequences. Functions $\text{ex}_S$ or their asymptotics have been determined for some matrices $S$ [7, 21, 24] and these results have found applications mostly in discrete geometry [2, 5, 6, 18] and also in the analysis of algorithms [20]. The Füredi–Hajnal conjecture states that $\text{ex}_P(n)$ is linear in $n$ whenever $P$ is a permutation matrix. Marcus and Tardos proved this conjecture by a surprisingly simple argument [15]. This implied the relatively long standing Stanley–Wilf conjecture by Klazar’s reduction [12]. An improved reduction yielding the upper bound $2^O(k \log k) n$ on the size of a set of $n$-permutations with a forbidden $k$-permutation was found by the first author [3].

We modify the question of Füredi and Hajnal and study the maximal number $p_k(n)$ of 1-entries in an $n \times n$ matrix such that no $(k+1)$-tuple of columns contains all $(k+1)$-permutation matrices. It can be easily shown that $p_2(n) \leq 4n$. Indeed, consider an $n \times n$ matrix with at least $4n+1$ 1-entries. Remove the highest and the lowest 1-entry in every column and find a row with three 1-entries in the resulting matrix. The three columns containing these 1-entries contain every 3-permutation matrix.

**Theorem 3.**

\[
2n\alpha(n) - O(n) \leq p_3(n) \leq O(n\alpha(n)),
\]

\[
p_{2t+2}(n) = n^{2(\alpha(n)^t) + O(\alpha(n)^{t-1})} \text{ for every } t \geq 1 \text{ and } \]

\[
n^{2(\alpha(n)^t) - O(\alpha(n)^{t-1})} \leq p_{2t+3}(n) \leq n^{2(\alpha(n)^t) \log_2(\alpha(n)) + O(\alpha(n)^t)} \text{ for every } t \geq 1.
\]

The upper bounds from Theorem 3 are proven as Corollary 8 in Section 2.1 and the lower bounds as Corollary 25 in Section 3.2.

Let $S$ and $T$ be sequences. We say that $S$ contains a pattern $T$ if $S$ contains a subsequence $T'$ isomorphic to $T$, that is, $T$ can be obtained from $T'$ by a one-to-one renaming of the symbols. A sequence $S$ over an alphabet $\Gamma$ is a Davenport–Schinzel sequence of order $s$ (a DS$(s)$-sequence for short) if no symbol appears on two consecutive positions and $S$ does not contain the pattern $abab\ldots ba$ of length $s+2$. These sequences were introduced by Davenport and Schinzel [4] and found numerous applications in computational and combinatorial geometry. More can be found in the book [23]. Let $\lambda_s(n)$ be the maximum length of a Davenport–Schinzel sequence over $n$ symbols. The following are the current best bounds on $\lambda_s(n)$.

\[
2n\alpha(n) - O(n) \leq \lambda_3(n) \leq 2n\alpha(n) + O\left(n\sqrt{\alpha(n)}\right),
\]

\[
n^{2(\alpha(n)^t) - O(\alpha(n)^{t-1})} \leq \lambda_{2t+2}(n) \leq n \cdot 2^{(\alpha(n)^t) + O(\alpha(n)^{t-1})} \text{ for } t \geq 1 \text{ and } \]

\[
n^{2(\alpha(n)^t) - O(\alpha(n)^{t-1})} \leq \lambda_{2t+3}(n) \leq n \cdot 2^{(\alpha(n)^t) \log_2(\alpha(n)) + O(\alpha(n)^t)} \text{ for } t \geq 1.
\]

The upper bound on $\lambda_3$ is by Klazar [11], the lower bounds on $\lambda_s$ for $s > 3$ are by Agarwal, Sharir and Shor [1] and all the other bounds were proved by Nivasch [17].
Our proofs are based on several results on Davenport–Schinzel sequences as well as on sequences with other forbidden patterns. The results on sequences that we use are mentioned in more detail in Sections 2.1, 2.2 and 3.1 where they are transformed into claims about matrices with forbidden patterns.

An \( s \)-partition of the rows of a matrix \( M \) is a partition of the interval of integers \( \{1, \ldots, m\} \) into \( s \) intervals \( \{1 = m_1, \ldots, m_2 - 1\}, \{m_2, \ldots, m_3 - 1\}, \ldots, \{m_s, \ldots, m = m_{s+1} - 1\} \). A matrix \( M \) contains a \( B \)-fat \((r,s)\)-formation if there exists an \( s \)-partition of the rows and an \( r \)-tuple of columns each of which has \( B \) 1-entries in each interval of rows. See Fig. [1] for an example of a 1-fat \((r,s)\)-formation.

In Section 2.2, we prove the following lemma, which is then used in the proof of Theorem 1 in Section 2.3, analogously to the use of Raz’s Technical Lemma [22].

Lemma 4. Let \( \beta_s(m) \) be the slowly growing functions defined in Section 2.2. For all positive integers \( m, n, s \) and \( B \), an \( m \times n \) matrix \( M \) with at least \( 2(\beta_s(m)+2)Bn \) 1-entries contains a \( B \)-fat \((\lfloor (n-1)B/(mc'_s)\rfloor, s)\)-formation, where \( c'_s \) is a constant depending only on \( s \).

The proof of the lemma is based on a proof of the upper bound on the number of symbols in the so-called formation-free sequences (see definition in Section 2.1) from the Nivasch’s paper [17].

All logarithms in this paper are in base 2.

2 Upper bounds

2.1 Numbers of 1-entries in matrices

A sequence \( S \) of length \( l \) over an alphabet \( \Gamma \) is a function \( S : [l] \rightarrow \Gamma \). An \((r,s)\)-formation is a sequence formed by \( s \) concatenated permutations of the same \( r \)-tuple of symbols. The permutations in a formation are its troops. A sequence \( S = (a_1, \ldots, a_l) \) is \( r \)-sparse if \( a_i \neq a_j \) whenever \( 0 < |i-j| < r \). An \((r,s)\)-formation-free sequence is a sequence that is \( r \)-sparse and contains no \((r,s)\)-formation as a subsequence. Let \( F_{r,s}(n) \) be the maximum length of an \((r,s)\)-formation-free sequence over \( n \) symbols. Formation-free sequences were first studied by Klazar [10].

To be able to use results on sequences for matrices, we define the matrix \( \rightarrow \) sequence transformation (MST). We read the given \( m \times n \) matrix \( M \) row by row ignoring 0-entries and writing, for each 1-entry, the index of the column in which it appears. The length of the resulting sequence \( \text{MST}(M) \) is equal to the number of 1-entries in \( M \) and the size of the alphabet is \( n \).

A block in a sequence is a contiguous subsequence containing only distinct symbols. Note that \( \text{MST}(M) \) can be decomposed into \( m \) or fewer blocks.

A set \( S \) of \( rs \) 1-entries forms an \((r,s)\)-formation in \( M \) if there exists an \( s \)-partition of the rows and an \( r \)-tuple of columns each of which has a 1-entry of \( S \) in every interval of rows of the partition. See Fig. [1]. In this and all other figures, circles and full circles represent the 1-entries and empty space represents the 0-entries. A matrix \( M \) is \((r,s)\)-formation-free if it contains no \((r,s)\)-formation.
Figure 1: A (3, 4)-formation on columns $j_1$, $j_2$ and $j_3$. Full circles represent the 1-entries of the formation. Empty circles represent 1-entries outside of this formation.

Observe that an $(r, s)$-formation in a matrix $M$ corresponds to an $(r, s)$-formation in $MST(M)$. The converse is not always true, because some two troops of the $(r, s)$-formation in $MST(M)$ may use symbols from the same block.

**Observation 5.** If $MST(M)$ of some matrix $M$ contains an $(sr, s)$-formation, then $M$ contains an $(r, s)$-formation.

**Proof.** Fix an $(sr, s)$-formation $F$ in $MST(M)$. A symbol of $F$ is bad if it is the last symbol of some troop, but not the last symbol of the whole formation. Since the symbols of $MST(M)$ are numbers, we can order the bad symbols into a sequence $b_1 \leq b_2 \leq \cdots \leq b_{s-1}$. Let $b_0 = 0$ and $b_s = n$, where $n$ is the number of columns of $M$. For every $i \in [s]$ let $\Gamma_i$ be the set of symbols $\{b_{i-1} + 1, \ldots, b_i\}$. Observe that for every $i \in [s]$ the columns corresponding to the symbols of $\Gamma_i$ contain a $(|\Gamma_i|, s)$-formation. The number of symbols of the formation is $rs$ and thus some $\Gamma_i$ has size at least $r$. □

Nivasch gives the following upper bound on the maximum length $F_{r,s}(n)$ of an $(r, s)$-formation-free sequence on $n$ symbols:

**Theorem 6.** ([17, Theorem 1.3]) For every $r \in \mathbb{N}$

$$F_{r,4}(n) \leq O(n\alpha(n)).$$

For every $r$ and every $s \geq 5$, letting $t := \lceil (s - 3)/2 \rceil$, we have

$$F_{r,s}(n) \leq n2^{(1/t)\alpha(n)^t \log(\alpha(n)) + O(\alpha(n)^{t-1})}$$

when $s$ is even and

$$F_{r,s}(n) \leq n2^{(1/t)\alpha(n)^{t-1} + O(\alpha(n)^{t-1})}$$

when $s$ is odd.

Let $p_k'(n)$ be the maximum number of 1-entries in an $(k + 1, k + 1)$-formation-free $n \times n$ matrix. Theorem 6 implies the following upper bounds on $p_k'(n)$.

**Lemma 7.**

$$p_3'(n) \leq O(n\alpha(n)).$$

For every fixed $k \geq 4$, letting $t := \lceil (k - 2)/2 \rceil$, we have

$$p_k'(n) \leq n2^{(1/t)\alpha(n)^t \log(\alpha(n)) + O(\alpha(n)^{t-1})}$$

when $k$ is odd and

$$p_k'(n) \leq n2^{(1/t)\alpha(n)^{t-1} + O(\alpha(n)^{t-1})}$$

when $k$ is even.
Proof. Take a \((k+1, k+1)\)-formation-free matrix \(M\). Then \(\text{MST}(M)\) does not contain a \(((k+1)^2, k+1)\)-formation by Observation 5.

The sequence \(\text{MST}(M) = a_1, a_2, \ldots, a_p\) can be made \((k+1)\)-sparse by removing at most \(kn\) occurrences of symbols. Indeed, whenever two occurrences \(a_i, a_j\) (where \(i < j\)) of the same symbol appear at distance at most \(k\), we remove symbols at positions \(i, i+1, \ldots, j-1\). This corresponds to shrinking \(M\) by at least one row and thus we stop after at most \(nk\) iterations, each of which removes at most \(k\) occurrences of symbols.

The resulting sequence is thus a \(((k+1)^2, k+1)\)-formation-free sequence of length differing by \(O(n)\) from the number of 1-entries of \(M\). The result then follows from Theorem 6.

This proves the upper bounds in Theorem 3 by observing that a matrix with a \((k+1, k+1)\)-formation contains every \((k+1)\)-permutation matrix.

Corollary 8. For every fixed \(k \geq 3\) if we let \(t := [(k-2)/2]\), then

\[
\begin{align*}
p_3(n) &\leq O(n \alpha(n)), \\
p_k(n) &\leq n2^{O(d)\alpha(n)^t+O(\alpha(n)^{t-1})} \quad \text{when } k \text{ is odd and greater than } 3 \text{ and} \\
p_k(n) &\leq n2^{O(d)\alpha(n)^t} \quad \text{when } k \text{ is even.}
\end{align*}
\]

Proof.

\[\blacksquare\]

2.2 Fat formations in matrices

A sequence \(S\) is an AFF\(_{r,s,k}(m)\)-sequence\(^1\) if it contains no \((r, s)\)-formation as a subsequence, can be decomposed into \(m\) or fewer blocks and each symbol of the sequence appears at least \(k\) times.

Let \(\Pi'_{r,s,k}(m)\) be the maximum number of symbols in an AFF\(_{r,s,k}(m)\)-sequence.

Let \(\alpha_d(m)\) be the \(d\)th function in the inverse Ackermann hierarchy. That is, \(\alpha_1(m) = \lceil m/2 \rceil\), \(\alpha_d(1) = 0\) for \(d \geq 2\) and \(\alpha_d(m) = 1 + \alpha_d(\alpha_{d-1}(m))\) for \(m, d \geq 2\).

The inverse Ackermann function is defined as \(\alpha(m) := \min \{k : \alpha_k(m) \leq 3\}\).

Nivasch defines a hierarchy of functions \(R_s(d)\), which we shift by 1 in the index. That is, our \(R_s(d)\) is the original \(R_{s-1}(d)\). We thus have the functions defined for \(s \geq 2\) and \(d \geq 2\). The values are \(R_2(d) = 2\), \(R_3(d) = 3\), \(R_4(d) = 2d + 1\), \(R_5(2) = 2s - 2 + 1\) and

\[
R_s(d) = 2(R_{s-1}(d) - 1) + (R_{s-2}(d) - 1)(R_s(d-1) - 3) + 1 \quad \text{when } s \geq 5 \text{ and } d \geq 3.
\]

For \(s \geq 5\), if we let \(t = [(s-3)/2]\), then \(R_s(d) = 2^{(1/\log d^{1/2}) \log d + O(d^{1/2})}\) if \(s\) is even and \(R_s(d) = 2^{(1/\log d^{1/2}) + O(d^{1/2})}\) when \(s\) is odd.

Lemma 9. ([17] Corollary 5.14]) For every \(d \geq 2\), \(s \geq 3\), \(r \geq 2\), \(m\) and \(k\) satisfying \(m \geq k \geq R_s(d)\) we have

\[
\Pi'_{r,s,k}(m) \leq c'_s r s \alpha_d(m)^{s-3},
\]

where \(c'_s\) is a constant depending only on \(s\).

---

\(^1\) AFF is an abbreviation for almost-formation-free.
The linear dependence of the upper bound on $r$ is not explicitly mentioned in [17], but can be revealed from the proof. In the base case, the dependence on $r$ is linear (Lemmas 5.9 and 5.10 in [17]) and in Recurrences 5.11 and 5.13, the right-hand side can be rewritten as $r$ times an expression not depending on $r$.

It was shown [13, 19] that doubling letters in the forbidden subsequence usually has small impact on the maximum length of a generalized DS-sequence (see the definition in [10]). Geneson [8] generalized the linear upper bound from the Füredi–Hajnal conjecture to forbidden double permutation matrices. We show a similar behavior of formation-free sequences and matrices. For $s \geq 2$, a set $S$ of $r(2s - 2)$ 1-entries forms a doubled $(r, s)$-formation in $M$ if there exists an $s$-partition of the rows and an $r$-tuple of columns each of which has one 1-entry of $S$ in the top and bottom interval of rows of the partition and two 1-entries in every other interval.

A matrix $M$ is doubled $(r, s)$-formation-free if it contains no doubled $(r, s)$-formation. A $\text{DFF}_{r,s,k}(m)$-matrix is a doubled $(r, s)$-formation-free matrix with $m$ rows and at least $k$ 1-entries in every column. Let $\Delta_{r,s,k}(m)$ be the maximum number of columns in a $\text{DFF}_{r,s,k}(m)$-matrix.

In Corollary 16 we show an analogue of Lemma 9 for doubled $(r, s)$-formation-free matrices. The proof follows the structure of the proof of Corollary 5.14 in [17]. The difference is that we formulate the claims for matrices and that some bounds get weaker, because we are forbidding doubled formations. First, we show some simple bounds on $\Delta_{r,s,k}(m)$. The case $d = 2$ of Corollary 16 is proved in Corollary 14 by Recurrence 13 and the remaining cases follow from Recurrence 15.

Corollary 16 will give a sequence of upper bounds on $\Delta_{r,s,k}(m)$. The bounds are indexed by $d$ and are improving as $d$ grows, but at the cost of higher requirements on $k$. For a given $m$, we then use the bound with $d = \alpha(m)$.

If $m < k$, no matrix with $m$ rows can have $k$ 1’s in every column.

**Observation 10.** For every $r, s, k, m$, if $m < k$, then

$$\Delta_{r,s,k}(m) = 0.$$  

**Observation 11.** For every $r, s, k, m$, if $k < 2s - 2$, then

$$\Delta_{r,s,k}(m) = \infty.$$  

Analogously to [17, Lemma 5.10], all the other values of $\Delta_{r,s,k}(m)$ are finite.

**Observation 12.** For every $r \geq 2$, $s \geq 2$ and $m \geq 2s - 2$

$$\Delta_{r,s,2s-2}(m) \leq (r - 1) \left( \begin{array}{c} m - s + 1 \\ s - 1 \end{array} \right) \leq rm^{s-1}.$$  

**Proof.** If each column in an $r$-tuple of columns has the same position of the 2nd, 4th, . . . , $(2s - 2)$nd 1-entry, then the first $2s - 2$ 1-entries from the columns form a doubled $(r, s)$-formation.

**Recurrence 13.** For every $r, k, m$ and $s \geq 3$

$$\Delta_{r,s,2k+1}(2m) \leq 2\Delta_{r,s,2k+1}(m) + 2\Delta_{r,s-1,k}(m).$$
Proof. As in the proof of [17, Recurrence 5.11], we cut the rows of a \( D F_{r,s,2k+1}(2m) \)-matrix into the upper \( m \) rows and the lower \( m \) rows. The \textit{local columns} are those with all 1-entries in the same half of rows. There are at most \( 2\Delta_{r,s,2k+1}(m) \) local columns. Columns that are not local are \textit{global}. Consider the submatrix \( M'_1 \) formed by the upper half of rows of global columns with at least half of their 1’s in the upper half of rows. Let \( M_1 \) be the matrix created from \( M'_1 \) by removing the lowest 1 in every column of \( M'_1 \). If \( M_1 \) contains a doubled \((r,s-1)\)-formation, then \( M \) contains a doubled \((r,s)\)-formation. Thus \( M_1 \) has at most \( \Delta_{r,s-1,k}(m) \) columns. A symmetric argument can be applied on the global columns with at least half of their 1’s in the lower half of rows.

\textbf{Corollary 14.} For every fixed \( s \geq 2 \) and for all integers \( r, k, m \) satisfying \( k \geq 2^s - 1 \) we have

\[ \Delta_{r,s,k}(m) \leq \bar{c}_s rm \log(m)^{s-2}, \]

where \( \bar{c}_s \) is a constant depending only on \( s \).

Proof. The proof proceeds by induction on \( s \) and \( m \). The base case of \( s = 2 \) follows from Observation [12] and the cases with \( m < k \) from Observation [10]. Recurrence [13] is used as the induction step.

\textbf{Recurrence 15.} For every nonnegative \( r, m, k_1, k_2, k_3, k_4 \) and \( t \) satisfying \( m > t \), \( k_3 \geq 3 \) and \( k_4 \geq k_3 \), if we let \( k = 2k_1 + (k_2 + 1)(k_3 - 3) + (k_4 - k_3) + 1 \), then

\[ \Delta_{r,s,k}(m) \leq (1 + \frac{m}{t}) \left( \Delta_{r,s,k}(t) + 2\Delta_{r,s-1,k_1}(t) + \Delta_{r,s-2,k_2}(t) \right) + \Pi'_{r,s,k_3} \left( 1 + \frac{m}{t} \right) + \Delta_{r,s,k_4} \left( 1 + \frac{m}{t} \right) \quad \text{for } s \geq 4 \]

\[ \Delta_{r,s,k}(m) \leq (1 + \frac{m}{t}) \left( \Delta_{r,s,k}(t) + 2\Delta_{r,s-1,k_1}(t) + r - 1 \right) + \Pi'_{r,s,k_3} \left( 1 + \frac{m}{t} \right) + \Delta_{r,s,k_4} \left( 1 + \frac{m}{t} \right) \quad \text{for } s = 3. \]

\textit{Proof.} Consider a \( D F_{r,s,k}(m) \)-matrix \( M \). We partition the rows of \( M \) into \( b := \lceil m/t \rceil \leq m/t + 1 \) layers \( L_1, \ldots, L_b \) of at most \( t \) consecutive rows each. A column is

- \textit{local} in layer \( L_i \) if all its 1’s appear in layer \( L_i \),
- \textit{top-concentrated} in layer \( L_i \) if it has at least \( k_1 + 1 \) 1’s in layer \( L_i \) and at least one 1-entry below \( L_i \),
- \textit{bottom-concentrated} in layer \( L_i \) if it has at least \( k_1 + 1 \) 1’s in layer \( L_i \) and at least one 1-entry above \( L_i \),
- \textit{middle-concentrated} in layer \( L_i \) if it has at least \( k_2 + 2 \) 1’s in layer \( L_i \) and at least one 1-entry above and one below layer \( L_i \),
- \textit{doubly-scattered} if it has at least two 1’s in at least \( k_3 \) layers,
- \textit{scattered} if it has a 1-entry in at least \( k_4 \) layers.
Every column falls into one of these categories, since otherwise, it would contain only at most $2k_1 + (k_2 + 1)(k_3 - 3) + (k_4 - k_3) \leq k - 1$-entries.

For each layer $L_i$, the number of columns local in $L_i$ is at most $\Delta_{r,s,k}(t)$. For every fixed $i$ we consider the columns that are top-concentrated in $L_i$ and take $M_i'$, the submatrix of $M$ defined by these columns and the rows of $L_i$. Let $M_i$ be obtained from $M_i'$ by removing the lowest 1-entry from every column. If $M_i$ contains a doubled $(r, s)$-formation, then $M$ contains a doubled $(r, s)$-formation. Thus there are at most $\Delta_{r,s-1,k_1}(t)$ columns top-concentrated in $L_i$. Similarly, there are at most $\Delta_{r,s-1,k_1}(t)$ columns bottom-concentrated in $L_i$. For $s \geq 4$, there are at most $\Delta_{r,s-2,k_2}(t)$ columns middle-concentrated in $L_i$. For $s = 3$, there are at most $r - 1$ columns middle concentrated in $L_i$, because an $r$-tuple of columns with two 1’s in layer $L_i$ and at least one 1 above and one below contains a doubled $(r, 3)$-formation.

To bound the number of doubly-scattered columns, we contract each layer into a single row. That is, we write 1 for every column containing at least two 1’s in the layer and 0 otherwise. If there is an $(r, s)$-formation on the contracted doubly-scattered columns, then $M$ contains a doubled $(r, s)$-formation. Thus, by Observation 5, there are at most $\Pi_{r,s,k_3}([m/t])$ doubly-scattered columns.

Similarly, the number of scattered columns is at most $\Delta_{r,s,k_4}([m/t])$. The only difference is that while contracting, we write 1 for the columns containing at least one 1 in the layer.

Similarly to the Nivasch’s functions $R_s(d)$, we define a hierarchy of functions $D_s(d)$, where $s \geq 1$ and $d \geq 2$, as follows: $D_1(d) = 0$, $D_2(d) = 2$, $D_s(2) = 2^{s-1} + 2^{s-2} - 1$ and when $s, d \geq 3$

$$D_s(d) = 2D_{s-1}(d) + (D_{s-2}(d) + 1)(R_s(d-1) - 3) + D_s(d-1) - R_s(d-1) + 1.$$  

Then

$$D_3(d) = 2d + 1, \quad D_4(d) \leq O(d^2), \quad D_5(d) \leq O(d^2d),$$

$$D_{2t+3}(d) \leq 2^{(1/t)d^2 + O(d^{-1})} \quad \text{and} \quad D_{2t+4}(d) \leq 2^{(1/t)d^3 \log(d) + O(d^2)} \quad \text{for } t \geq 1.$$

**Corollary 16.** For every $d \geq 2$, $s \geq 2$, $r \geq 2$, $m$ and $k$ satisfying $m \geq k \geq D_s(d)$ we have

$$\Delta_{r,s,k}(m) \leq c_s rm_a d(m)^{s-2},$$

where $c_s$ is a constant depending only on $s$.

**Proof.** The proof proceeds by induction on $d, s$ and $m$ similarly to the proof of [17, Corollary 4.12].

In the case $s = 2$, we apply Observation 12 and so the lemma holds with $c_2 = 1$. For every $s \geq 3$ let $m_0(s)$ be a constant such that

$$m \geq 1 + (6s)^s \lfloor \log_2(m) \rfloor s^2 \quad \text{for every } m \geq m_0(s).$$

Let $\hat{c}_1 = \hat{c}_2 = 1$ and for $s \geq 3$ we define $\hat{c}_s$ in the order of increasing $s$ as

$$\hat{c}_s := \max\{c'_s, \hat{c}_s, 9\hat{c}_{s-1}, 9\hat{c}_{s-2}, m_0(s)^{s-1}\},$$
where $\tilde{c}_s$ is the constant from Corollary 14 and $\tilde{c}_s'$ is the constant from Lemma 9.

For every $s \geq 3$ and $d \geq 2$, we define a function $\tilde{\alpha}_{d,s}$ by $\tilde{\alpha}_{2,s}(m) = [\log(m)]$, $\tilde{\alpha}_{d,s}(m) = 1$ if $m \leq m_0(s)$ and

$$\tilde{\alpha}_{d,s}(m) = 1 + \tilde{\alpha}_{d,s}(6s\tilde{\alpha}_{d-1,s}(m)^{s-2})$$

otherwise. Then $\tilde{\alpha}_{d,s}(m)$ is well defined and differs by at most an additive constant (depending on $s$) from the values of the $d$th inverse Ackermann function $\alpha_d(m)$ for all $s$, $d$ and $m$ (this can be shown similarly to [17, Appendix C]). The functions also satisfy $\tilde{\alpha}_{d,s}(m) \geq \tilde{\alpha}_{d,s-1}(m)$.

It will be then enough to prove

$$\Delta_{r,s,k}(m) \leq \tilde{c}_s rm\tilde{\alpha}_{d,s}(m)^{s-2}.$$

The case $d = 2$ follows from Corollary 14. The cases $m \leq m_0(s)$ follow from Observation 12.

Now $s \geq 3$, $d \geq 3$ and $m > m_0(s)$. We apply Recurrence 15 with:

$$k_1 = D_{s-1}(d), \quad k_2 = D_{s-2}(d), \quad k_3 = R_s(d - 1),$$

$$k_4 = D_s(d - 1), \quad k = D_s(d) \quad \text{and} \quad t = 6s\tilde{\alpha}_{d-1,s}(m)^{s-2}.$$

By the induction hypothesis,

$$2\Delta_{r,s-1,k_1}(t) \leq 2\Delta_{r,s-2,k_2}(t) \leq \frac{\tilde{c}_s}{3} t \tilde{\alpha}_{d,s}(m)^{s-3} \quad \text{when} \; s \geq 4,$$

$$2\Delta_{r,s-1,k_1}(t) + r - 1 \leq \frac{\tilde{c}_s}{3} t \tilde{\alpha}_{d,s}(m)^{s-3} \quad \text{when} \; s = 3,$$

$$\Delta_{r,s,k_4} \left(1 + \frac{m}{t}\right) \leq \tilde{c}_s \frac{2m}{t} \tilde{\alpha}_{d-1,s}(m)^{s-2} \leq r \frac{\tilde{c}_s}{3} m \leq r \frac{\tilde{c}_s}{3} m \tilde{\alpha}_{d,s}(m)^{s-3} \quad \text{for all} \; s \geq 3$$

and by Lemma 9,

$$\Pi'_{r,s,k_3} \left(1 + \frac{m}{t}\right) \leq \tilde{c}_s \frac{2m}{t} \tilde{\alpha}_{d-1,s}(m)^{s-3} \leq \frac{\tilde{c}_s}{3} m \leq r \frac{\tilde{c}_s}{3} m \tilde{\alpha}_{d,s}(m)^{s-3}.$$

Substituting into Recurrence 15 we get

$$\Delta_{r,s,k}(m) \leq \frac{m}{t} \Delta_{r,s,k}(t) + \Delta_{r,s,k}(t) + (m + t) r \frac{\tilde{c}_s}{3} \tilde{\alpha}_{d,s}(m)^{s-3} + \frac{4\tilde{c}_s}{9} r m \tilde{\alpha}_{d,s}(m)^{s-3} \leq \frac{m}{t} \Delta_{r,s,k}(t) + \frac{7\tilde{c}_s}{9} r m \tilde{\alpha}_{d,s}(m)^{s-3} + \Delta_{r,s,k}(t) + \frac{\tilde{c}_s}{3} r t \tilde{\alpha}_{d,s}(m)^{s-3}.$$

By Observation 12, $\Delta_{r,s,k}(t) \leq rt^{s-1} \leq r(6s\tilde{\alpha}_{d-1,s}(m)^{s-2})^{s-1}$, which is at most $rm$, because $m \geq m_0(s)$. So $\Delta_{r,s,k}(t) \leq \tilde{c}_s r m / 9$. Similarly $t \tilde{\alpha}_{d,s}(m)^{s-3} \leq m / 3$. Thus

$$\Delta_{r,s,k}(m) \leq \frac{m}{t} \Delta_{r,s,k}(t) + \frac{\tilde{c}_s}{9} r m \tilde{\alpha}_{d,s}(m)^{s-3} + \frac{\tilde{c}_s}{9} r m + \frac{\tilde{c}_s}{9} r m \tilde{\alpha}_{d,s}(m)^{s-3} \leq \frac{m}{t} \Delta_{r,s,k}(t) + \tilde{c}_s r m \tilde{\alpha}_{d,s}(m)^{s-3} \leq \frac{m}{t} \tilde{c}_s r t \tilde{\alpha}_{d,s}(t)^{s-2} + \tilde{c}_s r m \tilde{\alpha}_{d,s}(m)^{s-3}$$

by the induction hypothesis

$$\leq m \tilde{c}_s r \cdot ((\tilde{\alpha}_{d,s}(m) - 1)^{s-2} + \tilde{\alpha}_{d,s}(m)^{s-3}) \leq \tilde{c}_s r m \tilde{\alpha}_{d,s}(m)^{s-2}.$$

$\square$
Let $\beta_s(m) := D_s(\alpha(m))$.

**Corollary 17.** An $m \times n$ matrix $M$ with at least $\beta_s(m)$ 1-entries in every column contains an $\lfloor \left( (n-1)/(mc'_{s}) \right) \rfloor, s$)-formation, where $c'_{s}$ is a constant depending only on $s$.

**Proof.** Let $c'_{s} = c_s 3^{s-3}$, where $c_s$ is the constant from Corollary 16 and let $r = \lfloor (n-1)/(mc'_{s}) \rfloor$. If $M$ did not contain a doubled $(r, s)$-formation, it would, by Corollary 16 with $d = \alpha(m)$, have at most
\[
c_s r m \alpha(m) (m)^{s-3} \leq r m c_s 3^{s-3} = \lfloor (n-1)/(mc'_{s}) \rfloor m c'_s < n
\]
columns.

A set $S$ of $Brs$ 1-entries forms a $B$-fat $(r, s)$-formation in $M$ if there exists an $s$-partition of the rows and an $r$-tuple of columns each of which has $B$ 1-entries of $S$ in each interval.

A matrix $M$ is $B$-fat $(r, s)$-formation-free if it contains no $B$-fat $(r, s)$-formation.

**Proof of Lemma 4.** We transform the given matrix $M$ to a matrix $\overline{M}$ with the same number of 1-entries in every column using the idea from the proof of Lemma 4.1 from [17]. Let $v(q)$ be the number of 1-entries in a column $q$ of $M$. In every column $q$, we split the 1-entries into chunks of consecutive $(\beta_s(m)+2)$ 1’s. The last less than $(\beta_s(m)+2)B$ 1’s are discarded. Each of the chunks gets its own column with 1-entries in the rows where the 1-entries of the chunk lie. These columns form the matrix $\overline{M}$. Note that the order in which the columns are placed to $\overline{M}$ is not important. Because we discarded at most $(\beta_s(m)+2)Bn$ 1’s and every column of $\overline{M}$ has exactly $(\beta_s(m)+2)B$ 1’s, $\overline{M}$ has at least $n$ columns. Observe that for every $r$ and $s$ if $\overline{M}$ contains a $B$-fat $(r, s)$-formation, then so does $M$.

We consider only the first $n$ columns of $\overline{M}$. We also remove at most $B - 1$ rows so as to have the number of rows divisible by $B$. We still have at least $(\beta_s(m)+1)B$ 1’s in every column. In each column $q$, we select a set $S$ of 1’s such that none of them is among the first or the last $B - 1$ 1’s of $\overline{M}$ and there are at least $B - 1$ 1’s between every two 1’s of $S$. We take $S$ of size $\beta_s(m)$ and remove all the other 1’s in $q$. The rows of $\overline{M}$ are now grouped into intervals of rows $\{iB+1, \ldots, (i+1)B\}$. By the choice of $S$, every column contains at most one 1-entry in every interval. We obtain $\overline{M}$ by contracting each of the intervals of rows into a single row.

The matrix $\overline{M}$ has $\lfloor m/B \rfloor$ rows and $n$ columns, each of them having $\beta_s(m)$ 1’s. It thus contains a doubled $\lfloor \left( (n-1)B/(mc'_{s}) \right) \rfloor, s$)-formation by Corollary 17. By the choice of $S$, this implies that $\overline{M}$ and consequently $M$ contain a $B$-fat $\lfloor \left( (n-1)B/(mc'_{s}) \right) \rfloor, s$)-formation.

\[\Box\]

### 2.3 Sets of permutations with bounded VC-dimension

In this section we prove Theorem 11.

It will be more convenient for the proof to substitute the permutations by their corresponding permutation matrices. That is, we have a set $\mathcal{P}$ of $n$-permutation matrices
and for every \((k + 1)\)-tuple \((a_1, \ldots, a_{k+1})\) of columns, there is a forbidden \((k + 1)\)-permutation matrix \(S_{a_1,\ldots,a_{k+1}}\).

Let \(M_P\) be a \((0, 1)\)-matrix with 1-entries on the positions where at least one matrix from \(\mathcal{P}\) has a 1-entry.

Let \(|M|\) denote the number of 1-entries in a \((0, 1)\)-matrix \(M\) and let \(v(\mathcal{P}) = v(M_P) = |M_P|/n\). Similarly to the Raz’s proof of the exponential upper bound on \(r_2(n)\) \cite{22}, we will be removing matrices from \(\mathcal{P}\) in order to decrease \(v(\mathcal{P})\) below some threshold \(T(n)\).

When \(v(\mathcal{P}) \leq T(n)\), then \(|\mathcal{P}| \leq T(n)^n\) since the number of permutation matrices contained in \(M_P\) is bounded from above by the maximum of a product of \(n\) numbers with sum \(v(\mathcal{P})n\).

Let \(\gamma_k(n) = 4(\beta_{k+1}(n)+2)c_{k+1}'(k+1)!\), where \(c_k\) is the constant from Lemma \[4\]. Thus \(\gamma_k(n) \in O(\beta_{k+1}(n))\), where the constant hidden by the \(O\)-notation depends only on \(k\).

**Lemma 18.** Let \(\mathcal{P}\) be a set of \(n\)-permutation matrices with VC-dimension \(k\) such that \(v(\mathcal{P}) \geq 2\gamma_k(n)\). Then there is a set \(\mathcal{P}' \subset \mathcal{P}\) satisfying

\[
v(\mathcal{P}') \leq v(\mathcal{P}) - \frac{v(\mathcal{P})^2}{\gamma_k^2(n)n} \\
|\mathcal{P}'| \geq \frac{|\mathcal{P}|}{2v(\mathcal{P})^{2k}}.
\]

**Proof.** Let \(B := \left(\frac{v(\mathcal{P})}{(2(\beta_{k+1}+2))}\right)\). By Lemma \[4\] \(M_P\) contains a \(B\)-fat \([(B-1)/c_{k+1}', k + 1)\)-formation. Let \(C\) be the set of columns of the formation and let \(R = \{R_1, R_2, \ldots, R_{k+1}\}\) be the intervals of rows of the formation.

Consider some \(t\)-tuple \(Q = \{q_1, \ldots, q_t\}\) of columns from \(C\), where \(t \leq k + 1\). Let \(\mathcal{I}_Q\) be the set of all injective functions \(I : Q \to R\) assigning the intervals \(R_j\) to the columns \(q_j\). We say that a permutation matrix \(P\) obeys \(I \in \mathcal{I}_Q\) if for every \(i \in \{1, 2, \ldots, t\}\), the 1-entry of \(P\) in the column \(q_i\) lies in some row of \(I(q_i)\). For each \(I \in \mathcal{I}_Q\) let \(\mathcal{P}_I\) be the set of matrices \(P \in \mathcal{P}\) that obey \(I\).

The \(t\)-tuple \(Q\) of columns is said to be **criss-crossed** if

\[
\forall I \in \mathcal{I}_Q : |\mathcal{P}_I| \geq |\mathcal{P}|/v(\mathcal{P})^{2t}.
\]

Suppose that some \((k + 1)\)-tuple of columns from \(C\) is criss-crossed. Then every \((k + 1)\)-permutation appears as a restriction of some matrix from \(\mathcal{P}\) on the criss-crossed \((k + 1)\)-tuple of columns. Hence the VC-dimension of \(\mathcal{P}\) is at least \(k + 1\).

Consequently, there is some \(t\) such that \(0 \leq t \leq k\) and the largest criss-crossed set \(Q\) of columns from \(C\) has size \(t\). This means that for every column \(u\) outside \(Q\), we can find an injective function \(J_u \in \mathcal{I}_{Q \cup \{u\}}\) such that \(|\mathcal{P}_J_u| < |\mathcal{P}|/v(\mathcal{P})^{2t+2}\). On the other hand, if we restrict \(J_u\) on \(Q\), the resulting function \(I_u := J_u \upharpoonright Q\) satisfies \(|\mathcal{P}_I_u| \geq |\mathcal{P}|/v(\mathcal{P})^{2t}\). To each choice of \(u \in C \setminus Q\), we assign the function \(I_u \in \mathcal{I}_Q\) and the interval \(J_u(u)\) of rows.

Some function \(I \in \mathcal{I}_Q\) was then assigned to at least

\[
|C| - k \geq |C| - k \geq \frac{v(\mathcal{P})}{\gamma_k(n)}
\]

columns. Let \(\mathcal{T}_I\) be the set of some \([v(\mathcal{P})/\gamma_k(n)]\) columns that were assigned the function \(I\). For each column \(q_i \in Q\), we remove from \(M\) all 1-entries in the column \(q_i\) except those
that are in some of the rows of $I(q_i)$. This reduces the number of permutation matrices, but there are still at least $|P|/v(P)^{2t}$ of them. Then we remove from $M$ the 1-entries in each column $u \in T_I$ that lie in the set of rows $J_u(u)$. See Fig. 2. Thus we removed at least $B$ 1-entries from each of these columns. The removed 1-entries of each of these columns decreased the number of permutation matrices by at most $|P_J| \leq |P|/v(P)^{2t+2} + 2$.

![Figure 2: A criss-crossed set $Q$ of 3 columns and a set $T_I$ of columns $\{u_1, \ldots, u_5\}$, where $I(q_1) = R_4$, $I(q_2) = R_1$, $I(q_3) = R_2$, $J_{u_1}(u_1) = J_{u_2}(u_2) = R_3$ and $J_{u_3}(u_3) = J_{u_4}(u_4) = J_{u_5}(u_5) = R_5$. The 1-entries from the crossed rectangles are removed.](image)

Let $P' \subset P$ be the set of permutation matrices containing none of the removed 1-entries.

We have

$$|M_{P'}| \leq |M_P| - B|T_I| \leq nv(P) - \frac{v(P)^2}{\gamma_k^2(n)},$$

$$|P'| \geq \frac{|P|}{v(P)^{2t}} - \frac{|P||T_I|}{v(P)^{2t+2}} \geq \frac{|P|}{2v(P)^{2t}}.$$

Proof of Theorem 1. Let $P$ be a set of permutation matrices with VC-dimension $k$. We will bound its size by iteratively applying Lemma 18. Let $P_0 = P$ and for $j \geq 1$ let $P_j$ be the $P'$ given by the lemma applied on $P_{j-1}$.

The iterations are further grouped into phases. Let $\phi_0 := 0$. Phase $i$ ends after the first iteration $\phi_i$ after which $v(P_{\phi_i}) \leq v(P_{\phi_{i-1}})/2$. Let $v_i := v(P_{\phi_i})$. Then an iteration of phase $i$ is applied on a set $P$ of permutations satisfying

$$\frac{v_{i-1}}{2} \leq v(P) \leq v_{i-1} \quad (1)$$

Further, let

$$T := \gamma_k^2(n) \log(\gamma_k(n)). \quad (2)$$

We end after the first phase $l$ satisfying $v_l \leq 2T$. We thus have

$$|P_{\phi_l}| \leq (2T)^n. \quad (3)$$

13
For every \( i \geq 1 \) we have
\[
v_{i-1} \geq 2^iT. \tag{4}\]

We now count the number of iterations in phase \( i \). By Lemma 18 and (1), each of these iterations decreases \( v(P) \) by at least \( v_{i-1}^2/(4\gamma_k^2(n)n) \). Therefore the phase ends after at most \( [2\gamma_k^2(n)n/v_{i-1}] \leq 3\gamma_k^2(n)n/v_{i-1} \) iterations.

Consequently
\[
|P_{\phi_{i-1}}| \leq |P_{\phi_i}| \cdot \left(2v_{i-1}^2\right)^{3\gamma_k^2(n)n/v_{i-1}} \quad \text{by Lemma 18 and (1)}
\]
\[
\leq |P_{\phi_i}| \cdot 2^{(1+2k\log v_{i-1})3\gamma_k^2(n)n/v_{i-1}}
\]
\[
\leq |P_{\phi_i}| \cdot 2^{9k\gamma_k^2(n)n \log v_{i-1}/v_{i-1}}
\]

and
\[
|P| = |P_0| \leq |P_{\phi_i}| \cdot \prod_{i=0}^{l-1} 2^{9k\gamma_k^2(n)n \log v_i/v_i}
\]
\[
\leq |P_{\phi_l}| \cdot 2^{9k\gamma_k^2(n)n \sum_{i=1}^{l} k\log(2^iT)/(2^iT)} \quad \text{by (4) and since } \frac{\log(x)}{x} \text{ is decreasing on } [2T, \infty)
\]
\[
\leq (2T)^n \cdot 2^{9k\gamma_k^2(n)n(2+log T)/T} \quad \text{by (3)}
\]
\[
\leq (2\gamma_k^2(n) \log(\gamma_k(n)))^n \cdot 2^{45kn} \quad \text{by (2)}.
\]

Since \( \gamma_k(n) \in O(\beta_{k+1}(n)) \), we have
\[
r_3(n) \leq \left(O(\alpha(n)^4 \log(\alpha(n)))\right)^n, \quad r_4(n) \leq 2^{n \cdot (2\alpha(n)+3 \log(\alpha(n))+O(1))},
\]
\[
r_{2t+2}(n) \leq 2^n \cdot (2\alpha(n)^t + O(\alpha(n)^{t-1})) \quad \text{for } t \geq 1 \quad \text{and}
\]
\[
r_{2t+3}(n) \leq 2^n \cdot (2\alpha(n)^t \log(\alpha(n)) + O(\alpha(n)^{t+1})) \quad \text{for } t \geq 1.
\]

\[ \square \]

**Remark.** Let an \((n,m)\)-function be a total function \( f : [n] \to [m] \). We say that a set \( \mathcal{F} \) of \((n,m)\)-functions has \textit{VC-dimension with respect to permutations} (abbreviated as \( p\text{VC-dimension} \)) \( k \) if \( k \) is the largest integer such that the set of restrictions of the functions in \( \mathcal{F} \) to some \( k \)-tuple of elements from \([n]\) contains all \( k \)-permutations. Let \( r_k'(n,m) \) be the size of the largest set of \((n,m)\)-functions with \( p\text{VC-dimension} \) \( k \).

Observe that we never used in the proof the fact that the matrices in \( \mathcal{P} \) have exactly one 1-entry in every row. Thus the upper bound from Theorem 1 also holds with \( r_k'(n,n) \) in place of \( r_k(n) \).
3 Lower bounds

3.1 Matrices from sequences

Let $DS_s$ be the $s \times 2$ matrix with 1-entry in the $i$th row and $j$th column exactly when $i + j$ is odd. For example

$$DS_4 = \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix} \quad \text{and} \quad DS_5 = \begin{pmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{pmatrix}.$$

Based on a construction of Davenport–Schinzel sequences of order 3 and length $\Omega(n\alpha(n))$ by Hart and Sharir [9], Füredi and Hajnal [7] constructed $n \times n DS_4$-avoiding matrices $A_n$ with $\Omega(n\alpha(n))$ 1-entries.

We will use a different construction of $DS(s)$-sequences of orders $s = 3$ and all even $s \geq 4$ by Nivasch [17] that together with the following transformation will provide us with $DS_{s+1}$-avoiding matrices with the additional property of having the same number of 1-entries in every column.

Let $S$ be a sequence over $n$ symbols that can be partitioned into $m$ blocks. Recall that each block contains only distinct symbols. We number the symbols $1, \ldots, n$ in the increasing order of their first appearance. The sequence–matrix transformation of $S$, $SMT(S)$, is the $m \times n (0,1)$-matrix with a 1-entry in the $i$th row and $j$th column exactly if the $i$th block in the sequence contains the symbol $j$.

**Observation 19.** ([7]) If $S$ is a sequence avoiding the alternating pattern $aba\ldots$ of length $s + 2$, then $SMT(S)$ avoids $DS_{s+1}$.

**Proof.** If $SMT(S)$ contains $DS_{s+1}$, then $S$ contains the alternating sequence $ba\ldots$ of length $s + 1$ for some $a < b$. By the numbering of the symbols, the symbol $a$ appears before the first occurrence of $b$, therefore $S$ contains $aba\ldots$ of length $s + 2$ and thus $S$ is not a $DS(s)$-sequence.

**Lemma 20.** For every $n$ there exists an $n \times n DS_4$-avoiding matrix $M_n$ with at least $2\alpha(n) - O(1)$ 1-entries in every column.

**Proof.** Let $A_d(x)$ be the $d$th function of the Ackermann hierarchy. We refer the reader to the Nivasch’s paper [17] for the definition. Let $A(x) = A_x(3)$ be the Ackermann function.

In Section 6 of [17], Nivasch constructs for every $d, m \geq 1$ an $ababa$-free sequence $Z_d(m)$. We use the sequences $Z_d' = Z_d(8d + 4)$ which have the following properties:

- Each symbol appears exactly $2d + 1$ times.
- The sequence can be decomposed into blocks of average length at least $4d + 2$ (by [17, Lemma 6.2]).
- The number $N_d$ of symbols of the sequence is at most $A_d(8d + 4 + c)$ (by [17, Lemma 6.2]), where $c$ is an absolute constant.

15
Let $M_d$ be the number of blocks of $Z'_d$. By counting the length of $Z'_d$ in two ways, $(2d + 1)N_d \geq (4d + 2)M_d$ and thus $N_d \geq 2M_d$. By the analysis before Equation (35) in [17], there is some $d_0$ such that for $d \geq d_0$ we have

$$N_d \leq A_d(8d + 4 + c) \leq A_d(A(d + 1)) = A(d + 2) \quad \text{and so} \quad \alpha(N_d) \leq d + 2.$$ 

Then $SMT(Z'_d)$ is an $M_d \times N_d$ matrix with $2d + 1 \geq 2\alpha(N_d) - 3$ 1-entries in every column. By Observation 19, $SMT(Z'_d)$ avoids $DS_4$. We construct the matrix $M_{N_d}$ by adding empty rows to $SMT(Z'_d)$.

For values $n \geq N_d$ different from $N_d$, we proceed similarly to the method in Section 6 of [17]: We consider the largest $N_d$ smaller than $n$ and take $\lceil n/N_d \rceil$ copies of $SMT(Z'_d)$. We place the copies into a single matrix so that each copy has its own set of consecutive rows and columns. After removing at most half of the columns, we obtain a matrix with exactly $n$ columns and at most $n$ rows. The matrix has at least $2d + 1 \geq 2\alpha(N_d + 1) - 5 \geq 2\alpha(n) - 5$ 1-entries in every column. The construction of $M_n$ is then finished by adding empty rows to obtain a square matrix.

**Lemma 21.** For every $t \geq 1$ and $n$ there exists an $n \times n$ $DS_{2t+3}$-avoiding matrix with at least $2^{(1/t)\alpha(n) - O(\alpha(n)^{-1})}$ 1-entries in every column. In particular,

$$\exp_{DS_{2t+3}}(n) \geq n^{2^{(1/t)\alpha(n) - O(\alpha(n)^{-1})}}.$$ 

**Proof.** Let $s := 2t + 2$. Since $s$ is even and $s \geq 4$, we can use Nivasch’s construction [17, Section 7] of $DS(s)$-sequences $S_k^s(m)$ with parameters $k, m \geq 2$.

Let $\mu_s(k) := 2^{(s-2)/2}$. We use the sequences $S_{s,k}' = S_k^s(2\mu_s(k))$, which have the following properties:

- Each symbol of $S_{s,k}'$ appears exactly $\mu_s(k)$ times (by [17, Equation (47)]).
- The sequence can be decomposed into blocks of length $2\mu_s(k)$.
- For every $k \geq k_0(s)$, where $k_0(s)$ is a properly chosen constant, the number $N_{s,k}$ of symbols of the sequence satisfies $\alpha(N_{s,k}) \leq k + 3$ (by [17, Equations (50), (51)] and analysis similar to the one in the proof of Lemma 20).

Let $M_{s,k}$ be the number of blocks of $S_{s,k}'$. It satisfies $2M_{s,k} \leq N_{s,k}$.

The matrix $SMT(S_{s,k}')$ is a $DS_{s+k}$-avoiding $M_{s,k} \times N_{s,k}$ matrix with at least $\mu_{s,k}$ 1-entries in every column. For every $n \geq N_{s,k}$, we take the largest $k$ such that $n \geq N_{s,k}$ and proceed in the same way as in the proof of Lemma 20 with $SMT(S_{s,k}')$ in the place of $SMT(Z'_d)$. We have

$$k \geq \alpha(N_{s,k+1}) - 4 \geq \alpha(n) - 4$$

and so the number of 1-entries in every column of the resulting matrix is

$$\mu_s(k) = 2^{(s-2)/2} = 2^{(k)} \geq 2^{(1/t)\alpha(k)^{-1}} \geq 2^{(1/t)\alpha(n)^{-1}}.$$ 

$\square$
Remark. We could also use the construction of $DS_2$-avoiding matrices with $\Omega(n\alpha(n))$ 1-entries by Füredi and Hajnal [7]. The matrices do not have the same number of 1-entries in every column, but it can be shown that every column has at most constant multiple of the average number of 1-entries per column. This would be enough for our purposes.

The base case of the inductive construction in [7] needs a small fix. The matrices $M(s, 1)$ and $M(1, s)$ do not satisfy conditions imposed on them. This can be fixed for example by taking $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for $M(s, 1)$ and the matrix with the leftmost column full of 1-entries and with no 1-entries in the other columns for $M(1, s)$.

3.2 Numbers of 1-entries in matrices

A matrix is $k$-full if some $k$-tuple of its columns contains every $k$-permutation matrix. The fullness of a matrix $A$ is the largest $k$ such that $A$ is $k$-full. In this section we show a lower bound on the maximum number $p_k(n)$ of 1-entries in an $n \times n$ matrix with fullness $k$. This is achieved by showing that a $k$-full matrix contains the matrix $DS_k$ and applying the results from Section 3.1. We prove a slightly stronger statement that will be used in the next section.

Let $J_2 := (\bullet^* \bullet^*)$. For an $l$-permutation matrix $P$, we define the $J_2$-expansion of $P$, $P^{J_2}$, to be the $2l \times 2l$ permutation matrix created by substituting every 1-entry of $P$ by $J_2$ and every 0-entry by a $2 \times 2$ block full of 0-entries.

A pair of rows $2i, 2i+1$ of $P^{J_2}$ will be called contractible if the 1-entry in row $2i$ is to the left of the 1-entry in row $2i+1$. That is, when $\pi^{-1}(i) < \pi^{-1}(i+1)$, where $\pi$ is the permutation corresponding to $P$.

To contract a pair of rows means to replace them by a single row with 1-entries in the columns where at least one of the two original rows had a 1-entry.

Let an $(n, m)$-function be a total function $f : [n] \rightarrow [m]$. A function matrix is a $(0, 1)$-matrix with exactly one 1-entry in every column. Assigning to a function $f$ a function matrix $G_f$ with $G_f(i,j) = 1 \iff f(j) = i$ provides a bijection between $(n, m)$-functions and $m \times n$ function matrices.

The set of $J_2$-expansion flattenings of $P$ is the set $\mathcal{F}(P^{J_2})$ of function matrices that can be obtained from $P^{J_2}$ by contracting some pairs of contractible rows. Let $\mathcal{P}_l$ be the set of $l$-permutation matrices and let

$$\Phi(l) := \{ \mathcal{F}(P^{J_2}) : P \in \mathcal{P}_l \}.$$ 

For example

$$\Phi(2) = \left\{ \left\{ \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix} \right\}, \left\{ \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix} \right\}, \left\{ \begin{pmatrix} \bullet & \bullet \\ \bullet & \bullet \end{pmatrix} \right\} \right\}.$$ 

Lemma 22. If an $n \times 2l$ matrix $A$ contains one matrix from $\mathcal{F}_P^{J_2}$ for every $l$-permutation matrix $P$, then $A$ contains an occurrence of $DS_{2l}$ on columns $\{2i-1, 2i\}$ for some $i \in [l]$.

Proof. We proceed by induction on $l$. The case $l = 1$ is trivial since $\Phi(l) = \{ \{J_2\} \} = \{ \{DS_2\} \}$.

The $ith$ pair of columns of $A$ is the pair of columns $\{2i-1, 2i\}$. For each $i \leq l$ let $h_i$ be the smallest number such that the $ith$ pair of columns of $A$ contains $J_2$ on a subset of rows $\{1 \ldots h_i\}$. Let $t$ be the largest number satisfying $\forall i \ h_i \leq h_t$. 

17
Let $A^t$ be the $(n-h_t+1) \times 2(l-1)$ matrix obtained from $A$ by removing the columns of the $t$th pair, removing top $h_t-1$ rows and then changing all 1-entries among the first $2(t-1)$ entries in the first row to 0’s. See Fig. 3

Figure 3: Induction step in the proof of Lemma 22. In this example $t = 3$.

For every $P$ with the topmost 1-entry in column $t$, $A$ contains an occurrence of some $F \in \mathcal{F}(P_{J_2})$, that uses the two 1-entries of the topmost occurrence of $J_2$ on the $t$th pair of columns. These occurrences induce an occurrence of some matrix from every set $F \in \Phi(l-1)$ in $A^t$.

By the induction hypothesis, $A^t$ contains $DS_{2(l-1)}$ on some $i$th pair of columns. By the choice of $t$, this occurrence of $DS_{2(l-1)}$ in $A$ does not use any of the rows $\{1 \ldots h_i\}$. Thus we obtain an occurrence of $DS_{2l}$ in $A$.

For an $l$-permutation matrix $P$ and $i \leq 2l+1$ we define $P_{J_2}(i)$ to be the $(2l+1)$-permutation matrix that becomes $P_{J_2}$ after removing the lowest row and column $i$. Then $\mathcal{F}(P_{J_2}, i)$ is the set of function matrices that can be obtained from $P_{J_2}(i)$ by contracting some pairs of contractible rows. For example

$$\mathcal{F}((\cdot \cdot), 4) = \left\{ \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}, \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}, \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}, \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \right\}.$$ 

Let

$$\Phi(l, i) := \{ \mathcal{F}(P_{J_2}, i) : P \in \mathcal{P}_l \}.$$ 

Lemma 23. If an $n \times (2l+1)$ matrix $A$ contains one matrix from $\mathcal{F}(P_{J_2}, i)$ for every $l$-permutation matrix $P$ and every $i \in [2l+1]$, then $A$ contains $DS_{2l+1}$.

Proof. For each $i \leq 2l+1$ let $d_i$ be the row number of the lowest 1-entry in the $i$th column of $A$. Let $t$ be the smallest number satisfying $\forall i \ d_i \geq d_t$.

Let $A^t$ be the $d_t \times 2l$ matrix obtained from $A$ by removing the $t$th column, removing rows below the $d_t$th row and then changing all 1-entries in the last row to the right of the removed column to 0’s.

Then $A^t$ contains one matrix from every $F \in \Phi(l)$, therefore by Lemma 22 the matrix $A^t$ contains an occurrence of $DS_{2l}$. By the choice of $t$, the matrix $A$ contains $DS_{2l+1}$.

$\square$
Corollary 24. For every \( k \geq 1 \)
\[
p_k(n) \geq \text{ex}_{DS_{k+1}}(n).
\]

Proof. When \( k \) is even, the result follows from Lemma 23 since for every \( l \)-permutation matrix \( P \) and for every \( i \in [2l+1] \), the set \( \mathcal{F}(P^l, i) \) contains some \((2l+1)\)-permutation matrix. The result for \( k \) odd follows from Lemma 22.

The results of this section and of Section 3.1 prove the lower bounds in Theorem 3.

Corollary 25.
\[
\begin{align*}
p_3(n) &\geq 2n\alpha(n) - O(n), \\
p_k(n) &\geq n2^{(1/t)\alpha(n)t-O(\alpha(n)t^{-1})} \quad \text{for } k \geq 4,
\end{align*}
\]
where \( t := \lfloor (k-2)/2 \rfloor \).

Proof. The lower bound for \( k = 3 \) is by Lemma 20 and Corollary 24 and from Lemma 21 and Corollary 24 when \( k \) is even and \( k > 3 \). When \( k \) is odd and \( k \geq 5 \), we use \( p_k(n) \geq p_{k-1}(n) \).

3.3 Sets of permutations

Proof of Theorem 2 Given \( k \) and \( n \), we take the \( DS_{k+1} \)-avoiding \( n \times n \) matrix \( A_{k,n} \) from Lemma 20 if \( k = 3 \) or from Lemma 21 if \( k \geq 4 \) is even. Let \( \rho_k(n) \) be the number of 1-entries that \( A_{k,n} \) has in every column, that is \( \rho_2(n) = 2\alpha(n) - O(1) \) and for \( t \geq 1 \)
\[
\rho_{2t+2}(n) = 2^{(1/t)\alpha(n)t-O(\alpha(n)t^{-1})}.
\]

From \( A_{k,n} \) we construct a set of \( \rho_k(n)^n \) \( n \times n \) function matrices by choosing some 1-entry from each column. Then we remove all empty rows, which can make some originally different function matrices identical. However, the resulting set \( \mathcal{H} \) has size at least \( \rho_k(n)^n/2^n \) as there are at most \( 2^n \) distinct ways to enlarge a function matrix by adding empty rows to a matrix with \( n \) rows.

The last step is inflating the rows of the function matrices in \( \mathcal{H} \) into diagonal matrices to obtain a set \( Q \) of \( n \)-permutation matrices. That is, for every \( H \in \mathcal{H} \), we order the 1-entries primarily by the rows from top to bottom and secondarily from left to right. The permutation matrix \( Q \) will have 1-entries on those positions \((i, j)\) such that \( H \) has its \( i \)th 1-entry in column \( j \). The reverse process consists of contracting intervals of rows of a permutation matrix \( Q \) and we have at most \( 2^n \) possibilities how to choose the intervals. Thus every permutation matrix can be created by expanding at most \( 2^n \) different function matrices.

The size of the set \( Q \) is
\[
|Q| \geq \left( \frac{\rho_k(n)^n}{2^n 2^n} \right)^n = \left( \frac{\rho_k(n)}{4} \right)^n.
\]

It remains to show that the VC-dimension of \( Q \) is at most \( (k+1) \). We assume for contradiction that for some \((k+1)\)-tuple \( C \) of columns and every \((k+1)\)-permutation matrix \( R \) there exists \( Q \in Q \) that contains \( R \) on \( C \).
Consider some permutation matrix $Q \in \mathcal{Q}$ and let $H \in \mathcal{H}$ be the function matrix from which $Q$ was created. The matrix $H$ can thus be constructed from $Q$ by contracting some intervals of rows such that the restriction of $Q$ on each of these intervals of rows is a diagonal matrix. So the only change that these contractions can make on an occurrence of $P^{J_2}$ in $Q$ is that some pairs of its contractible rows can be contracted. Thus an occurrence of $P^{J_2}$ in $Q$ on the set $C$ of columns can only be created from an occurrence of some $F \in \mathcal{F}(P^{J_2})$ on $C$ in $H$ and in $A_{k,n}$ as well. Similarly, an occurrence of $P^{J_2}(i)$ on $C$ in $Q$ implies an occurrence of some matrix from $\mathcal{F}(P^{J_2}, i)$ on $C$ in $H$ and $A_{k,n}$. See Fig. 4

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure4}
\caption{Expansion of an occurrence of a matrix from $\mathcal{F}(P^{J_2})$.}
\end{figure}

Therefore for $k \geq 4$ even and for every $(k/2)$-permutation matrix $P$ and every $i \in [k + 1]$, $A_{k,n}$ contains on $C$ an occurrence of some matrix from $\mathcal{F}(P^{J_2}, i)$. Thus, by Lemma 23, $A_{k,n}$ contains $DS_{k+1}$, a contradiction.

Similarly if $k = 3$, we get a contradiction by Lemma 22.

\begin{acknowledgements}
We thank Ottfried Cheong, Martin Klazar and Pavel Valtr for inspiring discussions.
\end{acknowledgements}

\begin{references}
\end{references}


