Combinatorics, Probability and Computing (1999) 8, 417–427. Printed in the United Kingdom
© 1999 Cambridge University Press

Sparse 0−1 Matrices and Forbidden Hypergraphs*

CLAUDIA BERTRAM-KRETZBERG,
THOMAS HOFMEISTER and HANNO LEFMANN†

Lehrstuhl Informatik 2, Universität Dortmund, D–44221 Dortmund, Germany
(e-mail: {bertram,hofmeist,lefmann}@Ls2.cs.uni-dortmund.de)

Received 20 January 1998; revised 19 January 1999

We consider the problem of determining the maximum number $N(m,k,r)$ of columns of a 0−1 matrix with m rows and exactly r ones in each column such that every k columns are linearly independent over $\mathbb{Z}_2$. For fixed integers $k \geqslant 4$ and $r \geqslant 2$, where k is even and $\gcd(k-1,r)=1$, we prove the lower bound $N(m,k,r) = \Omega(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}})$. This improves on earlier results from [14] by a factor $\Theta((\ln m)^{\frac{1}{k-1}})$. Moreover, we describe a polynomial time algorithm achieving this new lower bound.

1. Introduction

We consider matrices M with entries 0 and 1 with *k-wise independent* columns, that is, each set of k column vectors of M is linearly independent over $\mathbb{Z}_2 = \{0,1\}$. We call such a matrix a (k,r) *matrix* if, in addition, it has the property that each column contains exactly r ones. Let $N(m,k,r)$ denote the maximum number of columns a (k,r) matrix with m rows can have. Some special cases of this problem, for instance estimates on the growth of $N(m,k,2)$, and similar questions have been studied in combinatorics and graph theory with respect to forbidden configurations, like cycles in graphs, and there is an extensive literature on this topic (see [10]). Specifically, considering incidence matrices of graphs shows that $N(m,k,2)$ is equal to the maximum number of edges in a graph on m vertices without cycles of length $3,\ldots,k$, since in graphs, the minimal dependent configurations correspond to cycles. First estimates on $N(m,k,r)$ for arbitrary values of m,k,r, where k

* This research was supported by the Deutsche Forschungsgemeinschaft as part of the Collaborative Research Center 'Computational Intelligence' (SFB 531).
This paper was submitted to the Special Issue (Volume 8, Number 4) devoted to papers from the Oberwolfach meeting on *Random Graphs and Combinatorial Structures*, but was inadvertently omitted from that Issue.
† Part of this research was done during the author's visit to the Institut für Theoretische Informatik, Medizinische Universität Lübeck.

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

and r are fixed, were given in [14]. In this paper we improve on the lower bounds for $N(m,k,r)$ given there.

There is also a connection between the matrices considered in this paper and so-called *parity check matrices*. Recall that, for a linear code $C \subseteq \mathbb{Z}_2^n$, an $m \times n$ matrix M over $\mathbb{Z}_2$ is a parity check matrix for C if and only if for every $\underline{x} \in \mathbb{Z}_2^n$ we have that $\underline{x} \in C \Leftrightarrow M \cdot \underline{x} = \underline{0}$. The connection with the problem considered here is that a linear code C with parity check matrix M has minimum distance at least $k+1$ if and only if any k columns of M are linearly independent (see, for instance, Lemma 8.14 in [15]).

If there is no restriction on the number of ones per column, then a lower bound on the maximum number of columns of 0–1 matrices with k-wise independent columns is given by the Gilbert–Varshamov bound.

Besides their importance in coding theory, one further application for parity check matrices lies in the area of derandomization. They can be used for constructing small sample spaces preserving some limited independence. Also related are recent results of Sipser and Spielman [19, 21] who, motivated by the PCP theorem (see [22]), constructed asymptotically good codes using expander graphs. The parity check matrices of these codes are also sparse; however, not only each column but also each row has only a constant number of nonzero entries.

In [14], the function $N(m,k,r)$ was defined in a slightly different way: namely, each column was allowed to have at most r ones, but this does not change the growth rate asymptotically.

The following lower bound on $N(m,k,r)$ was shown in [14] by a probabilistic argument: for fixed integers $k \geqslant 2$ and $r \geqslant 2$,

$$N(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}}\right). \tag{1.1}$$

Moreover, [14] also provides an algorithm that computes (k,r) matrices achieving this lower bound. Its running time is $O(m^{kr/2})$, which is polynomial if k and r are fixed. Indeed, for certain values of the parameters k and r, the lower bound (1.1) is asymptotically sharp, that is, $N(m,k,r) = \Theta\left(m^{\frac{kr}{2(k-1)}}\right)$ (see [14]).

Kohayakawa, Kreuter and Steger [12] showed the following: for fixed $k \geqslant 2$, we have $N(m,2k,2) = \Omega(m^{2k/(2k-1)} \cdot (\ln m)^{1/(2k-1)})$. For this special case $r = 2$, there are better constructions known from the work of Lubotzky, Phillips and Sarnak [16] and Margulis [17]. However, the intention in [12] was to study Turán numbers in the random situation. The constructions from [16] and [17] use algebraic techniques and yield the so-called Ramanujan graphs, which are graphs on m vertices with at least $\Omega\left(m^{(3k+5)/(3k+3)}\right)$ edges which do not contain any cycle of length smaller than $2k+1$, that is, $N(m,2k,2) = \Omega\left(m^{(3k+5)/(3k+3)}\right)$. Recently Lazebnik, Ustimenko and Woldar [13] showed that $N(m,2k,2) = \Omega\left(m^{(3k-1+\beta)/(3k-3+\beta)}\right)$ with $\beta = 0$ if k is odd and $\beta = 1$ otherwise. Here, we do not focus on the case $r = 2$: we consider the case of arbitrary positive integers k and r.

As usual, let $\gcd(s,t)$ be the greatest common divisor of s and t. We improve the general lower bound (1.1) by the factor $\Theta((\ln m)^{1/(k-1)})$ if $k \geqslant 4$ is even and $\gcd(k-1,r) = 1$, that

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

is, we show that

$$N(m,k,r) = \Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\right). \tag{1.2}$$

The case of odd integers k can easily be reduced to the case of even integers since $1/2 \leqslant N(m,2k+1,r)/N(m,2k,r) \leqslant 1$ (see Section 3).

It seems surprising that our arguments do not work if $\gcd(k-1,r) > 1$. However, one cannot always improve the lower bound (1.1) by a logarithmic factor, as matching upper bounds sometimes do exist. As an example, consider that from the work of Frankl and Füredi [8] on union-free families of sets (compare also [14]) the following upper bound for $k = 4$ is known:

$$N(m,4,r) = O(m^{\lceil 4r/3\rceil/2}),$$

while (1.1) yields the lower bound $N(m,4,r) = \Omega(m^{2r/3})$. For $r \equiv 0 \bmod 3$, lower and upper bounds match up to constant factors. On the other hand, for, say, $r \equiv 1 \bmod 3$, the upper bound is only of the order $O(m^{2r/3+1/3})$. Our lower bound here shows $N(m,4,r) = \Omega(m^{2r/3} \cdot (\ln m)^{1/3})$ for $r \not\equiv 0 \bmod 3$.

The proof of our lower bound first transforms the problem into a problem on hypergraphs and then applies some probabilistic arguments. We also show that derandomization and recent results from [9] and [6] can be used to obtain a polynomial time algorithm for constructing (k,r) matrices of sizes as guaranteed by (1.2).

Upper bounds

For the existing bounds concerning the special case $r = 2$, we refer to [10]. We remark that matching lower and upper bounds are known only for $k \in \{4,6,10\}$: that is, for these values of k it is known that $N(m,k,2) = \Theta(m^{(k+2)/k})$.

Concerning upper bounds on $N(m,k,r)$ for arbitrary values of k and r, only the following recursions are known from [14]: namely, for fixed integers $k \geqslant 4$ and $r \geqslant 2$, where k is even, and $0 \leqslant s < r$ it holds that

$$\begin{aligned} N(m,k,r) &= O\left(N(m,k/2,2r-2s) + m^s\right), \quad \text{and} \\ N(m,k,r) &= O\left(\sqrt{m^s \cdot N(m,k/2,2r-2s)} + m^s\right). \end{aligned}$$

For $k = 2^i$ and $i \in \mathbb{N}$ we have that $N(m,k,r) = O\left(m^{\lceil kr/(k-1)\rceil/2}\right)$; hence, for $\gcd(k-1,r) = k-1$ we have $N(m,k,r) = \Theta\left(m^{\frac{kr}{2(k-1)}}\right)$, which yields other examples for which upper and lower bounds (1.1) agree.

2. Basic definitions

A *hypergraph* $\mathscr{G} = (V,\mathscr{E})$ consists of a set V of vertices and a set $\mathscr{E}$ of edges where every edge $E \in \mathscr{E}$ is a subset of V. A hypergraph is called *k-uniform* if all of its edges have cardinality k.

In our arguments 2-*cycles* are crucial, that is, pairs $\{E,E'\}$ of distinct edges $E,E' \in \mathscr{E}$ which satisfy $|E \cap E'| \geqslant 2$. A 2-cycle $\{E,E'\}$ is called $(2,j)$-*cycle* if $|E \cap E'| = j$. The *average degree* of a k-uniform hypergraph $(V,\mathscr{E})$ is defined by $d_{\text{avg}} = k \cdot |\mathscr{E}|/|V|$. It is more convenient to work with the parameter t defined by $t^{k-1} = d_{\text{avg}}$.

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

Given a subset $V_1 \subseteq V$ of the vertex set, the *induced* subhypergraph on V_1 has edge set $\{E \mid E \in \mathscr{E} \text{ and } E \subseteq V_1\}$. An *independent set* is a subset $I \subseteq V$ of the vertex set such that the induced subhypergraph on I contains no edge. The *independence number* $\alpha(\mathscr{G})$ of a hypergraph $\mathscr{G}$ is the size of a largest independent set.

3. Lower bounds on $N(m,k,r)$

It is easy to see that $N(m,k,1) = m$ and $N(m,2,r) = \binom{m}{r}$ as no column can occur more than once. For the asymptotic evaluation of $N(m,k,r)$, it suffices to consider even dependencies only, since the following inequalities hold:

$$N(m,2k+1,r) \begin{cases} \leqslant & N(m,2k,r), \\ \geqslant & 1/2 \cdot N(m,2k,r). \end{cases} \tag{3.1}$$

The first inequality follows from monotonicity; the second one was observed in [14] and can easily be proved as follows. One can show that every $m \times n$ matrix M with entries from $\{0,1\}$ and without the zero column contains an $m \times \lceil n/2 \rceil$ submatrix M' such that no odd number of columns of M' adds to zero. Namely, choose one of the 2^m subsets S of the rows according to the uniform distribution. The expected number of columns of M which have an odd number of ones in S is $n/2$; hence there is a subset S^* of the rows and a subset C of the column vectors with $|C| \geqslant \lceil n/2 \rceil$ such that all columns in C have an odd number of ones in S^*. The submatrix M' of M consisting only of the columns from C has the property that no odd number of columns adds to zero.

In the proof of our main result, we will use the following extension (see [7]) of a result of Ajtai, Komlós, Pintz, Spencer and Szemerédi [1].

Theorem 3.1. *Let $\mathscr{G}$ be a k-uniform hypergraph with $k \geqslant 3$. If $\mathscr{G}$ has n vertices, has average degree at most t^{k-1}, and has no 2-cycles, then*

$$\alpha(\mathscr{G}) = \Omega\Big(\frac{n}{t} \cdot (\ln t)^{\frac{1}{k-1}}\Big).$$

This result and variants of it (in particular the result from [1]) have been applied various times in recent years, for example in [2], [4], [18], [20] (see [6] for further references). Here, we obtain another application of Theorem 3.1, where we have to use fine-tuned arguments in order to keep the theorem applicable. Specifically, we prove the following new lower bound on $N(m,k,r)$.

Theorem 3.2. *Let $k \geqslant 4$ and $r \geqslant 2$ be fixed integers where k is even and* $\gcd(k-1,r) = 1$. *Then*

$$\begin{aligned} N(m,k,r) &= \Omega\Big(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\Big) \quad \textit{as well as} \\ N(m,k+1,r) &= \Omega\Big(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\Big). \end{aligned}$$

We use the following notation. Let C_m^r be the set of all 0–1 column vectors of length m with exactly r ones, so that $|C_m^r| = \binom{m}{r}$.

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

Proof. The second statement follows from the first and inequality (3.1). Let $\oplus$ be the componentwise addition in $\mathbb{Z}_2^m$. We form a hypergraph $\mathscr{G} = (V, \mathscr{E}_3 \cup \cdots \cup \mathscr{E}_k)$ with vertex set $V = C_m^r$ and i-element edges $\{a_1, \ldots, a_i\} \in \mathscr{E}_i$ if and only if $a_1 \oplus \cdots \oplus a_i = 0$ where $i = 3, \ldots, k$. By this definition, an independent set in $\mathscr{G}$ corresponds to a set of column vectors where each subset of k vectors is linearly independent. Thus, it is our aim to find a large independent set.

First we give an upper bound on the number $|\mathscr{E}_i|$ of i-element edges, $i = 3, \ldots, k$. Consider an edge E from $\mathscr{E}_i$ consisting of the vertices (column vectors) $a_1, \ldots, a_i \in C_m^r$. We can view this edge as an $m \times i$ matrix $M(E)$. Since $a_1 \oplus \cdots \oplus a_i = 0$, the number ir of ones in $M(E)$ is even and the number of ones in each row of $M(E)$ is also even. There are at most $(\binom{i}{2} \cdot m)^{ir/2}$ such matrices, which can be seen by considering the possible positions of pairs of ones. Thus, for $i = 3, \ldots, k$ we have for some positive constant $c_i = c_i(k, r)$ that

$$|\mathscr{E}_i| \leqslant c_i \cdot m^{ir/2}. \tag{3.2}$$

For the moment, let us neglect, with foresight, the edges from $\mathscr{E}_3 \cup \cdots \cup \mathscr{E}_{k-1}$ and consider only the k-uniform hypergraph $\mathscr{G}_k = (V, \mathscr{E}_k)$. We will first count (for $j = 2, \ldots, k-1$) the number of $(2, j)$-cycles in $\mathscr{G}_k$. We denote this number by $s_j(\mathscr{G}_k)$.

Consider a fixed j-element set $J = \{a_1, \ldots, a_j\}$ of column vectors ($j < k$) and view it as an $m \times j$ matrix $M(J)$. Let $p(J)$ be the number of rows of $M(J)$ that contain at least one entry 1. Let $p_{\text{odd}}(J)$, (respectively, $p_{\text{even}}(J)$) be the number of rows of $M(J)$ containing an odd (respectively, even) nonzero number of ones. With $p_{\text{even}}(J) + p_{\text{odd}}(J) = p(J)$ it follows that

$$2 \cdot p(J) - p_{\text{odd}}(J) = 2 \cdot p_{\text{even}}(J) + p_{\text{odd}}(J) \leqslant j \cdot r.$$

Also note that $p(J)$ is bounded by a constant, since $p(J) \leqslant k \cdot r$. Let $a(J)$ be the number of $(k-j)$-element subsets $\{b_1, \ldots, b_{k-j}\}$ from $C_m^r \setminus \{a_1, \ldots, a_j\}$ such that $E = \{a_1, \ldots, a_j, b_1, \ldots, b_{k-j}\} \in \mathscr{E}_k$, that is, $a_1 \oplus \cdots \oplus a_j \oplus b_1 \oplus \cdots \oplus b_{k-j} = 0$, so that the $m \times k$ matrix $M(E)$ consisting of these columns has an even number of ones in each row. Thus, for some positive constant $C_0 = C_0(k, r)$, there are

$$a(J) \leqslant (k-j)^{p_{\text{odd}}(J)} \cdot \left(m \cdot \binom{k-j}{2} \right)^{\frac{(k-j)r - p_{\text{odd}}(J)}{2}} \leqslant C_0 \cdot m^{\frac{kr - 2p(J)}{2}}$$

many ways to choose the columns $b_1, \ldots, b_{k-j}$. This holds since we can construct the matrix $M(E)$ from $M(J)$ by adding single ones into the rows which have an odd number of ones and then adding the remaining ones in pairs per row.

We now count quite carefully the number of $(2, j)$-cycles in $\mathscr{G}_k$.

For $j = 2, \ldots, k-1$ and $u = r+1, \ldots, jr$, let $s_{j,u}(\mathscr{G}_k)$ be the number of $(2, j)$-cycles $\{E_1, E_2\}$ in $\mathscr{G}_k$ such that $p(E_1 \cap E_2) = u$, that is, with u nonzero rows in the matrix $M(E_1 \cap E_2)$. Clearly, we have $s_j(\mathscr{G}_k) = \sum_{u=r+1}^{jr} s_{j,u}(\mathscr{G}_k)$.

For a fixed j-element subset $J \in [C_m^r]^j$ of column vectors, there are at most $\binom{a(J)}{2}$ pairs $\{J_1, J_2\}$ of $(k-j)$-element sets of column vectors such that $E_1 = J \cup J_1$ and $E_2 = J \cup J_2$ form a $(2, j)$-cycle in $\mathscr{G}_k$ with $E_1 \cap E_2 = J$. Hence, for some positive constant $C_1 = C_1(k, r)$,

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

the following holds:

$$\begin{aligned} s_{j,u}(\mathscr{G}_k) &\leqslant \sum_{J\in[C_m^r]^j:p(J)=u} \binom{a(J)}{2} \\ &\leqslant C_1 \cdot \sum_{J\in[C_m^r]^j:p(J)=u} m^{kr-2u}. \end{aligned} \tag{3.3}$$

For some subset V_1 of the vertex set V, let $p_{j,u}(V_1)$ denote the number of j-element subsets $J \in [V_1]^j$ of column vectors with $p(J) = u$. Then, for $j < k$, the value of $p_{j,u}(V)$ can be bounded from above by $\binom{m}{u}\cdot\binom{u}{r}^j \leqslant C_2' \cdot m^u$ for some positive constant $C_2' = C_2'(k,r)$, and inequality (3.3) becomes

$$s_{j,u}(\mathscr{G}_k) \leqslant C_1' \cdot m^{kr-u}. \tag{3.4}$$

The average degree $d_{\text{avg}} = t^{k-1}$ of the k-uniform hypergraph $\mathscr{G}_k = (V, \mathscr{E}_k)$ satisfies $t^{k-1} \leqslant k \cdot c_k \cdot m^{kr/2}/\binom{m}{r}$; hence, for some positive constant $C_k' = C_k'(k,r)$, we have

$$t \leqslant t_+ := C_k' \cdot m^{\frac{(k-2)r}{2(k-1)}}.$$

To apply Theorem 3.1, we want to find a subhypergraph of $\mathscr{G}$ without 2-cycles among the k-element edges and without i-element edges, $i = 3,\ldots,k-1$. To obtain such a subhypergraph, we choose in $\mathscr{G}$ at random vertices (column vectors) independently of each other with probability $p = t_+^{-1} \cdot m^{\delta}$, where δ is a small positive constant which will be specified later. Vertices not chosen are removed from the hypergraph. (We remark that the choice $p = t_+^{-1}$ would lead to the lower bound (1.1).)

Let V^* be the resulting random subset of V and let $\mathscr{G}^* = (V^*, \mathscr{E}_3^* \cup \cdots \cup \mathscr{E}_k^*)$ be the induced random subhypergraph of $\mathscr{G}$. We consider the expected values $\mathbb{E}[\,\cdot\,]$ of the parameters of the hypergraph $\mathscr{G}^*$. For some positive constant c,

$$\mathbb{E}[|V^*|] = p \cdot \binom{m}{r} \geqslant c \cdot m^{\frac{kr}{2(k-1)}+\delta}. \tag{3.5}$$

The expected number of i-element edges, $i = 3,\ldots,k$, satisfies by (3.2) for some positive constants $D_i = D_i(k,r)$ that

$$\begin{aligned} \mathbb{E}[|\mathscr{E}_i^*|] &\leqslant p^i \cdot c_i \cdot m^{\frac{ir}{2}} \\ &\leqslant D_i \cdot m^{\frac{ir}{2}-\frac{(k-2)ir}{2(k-1)}+\delta\cdot i} \\ &= D_i \cdot m^{\frac{ir}{2(k-1)}+\delta\cdot i}. \end{aligned} \tag{3.6}$$

Since $p_{j,u}(V) \leqslant C_2' \cdot m^u$, we have for some positive constant $C' = C'(k,r)$ that, for $j = 2,\ldots,k-1$ and $u = r+1,\ldots,jr$, the expected number of j-element subsets J of V^* with $p(J) = u$ satisfies

$$\begin{aligned} \mathbb{E}[p_{j,u}(V^*)] &= p^j \cdot p_{j,u}(V) \\ &\leqslant C_2' \cdot p^j \cdot m^u \\ &\leqslant C' \cdot m^{u-\frac{(k-2)jr}{2(k-1)}+\delta\cdot j}. \end{aligned} \tag{3.7}$$

For $j = 2,\ldots,k-1$ and $u = r+1,\ldots,jr$, the expected number of $(2,j)$-cycles $\{E_1, E_2\}$ in

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

the k-uniform hypergraph $\mathscr{G}_k^* = (V^*, \mathscr{E}_k^*)$ with $p(E_1 \cap E_2) = u$ satisfies by (3.4):

$$\begin{aligned} \mathbb{E}[s_{j,u}(\mathscr{G}_k^*)] &= p^{2k-j} \cdot s_{j,u}(\mathscr{G}_k) \\ &\leqslant C_1' \cdot p^{2k-j} \cdot m^{kr-u} \\ &\leqslant C_2 \cdot m^{\frac{(k-2)j+2k}{2(k-1)} \cdot r - u + \delta \cdot (2k-j)}. \end{aligned} \tag{3.8}$$

By Chebyshev's and Markov's inequalities there exists an induced subhypergraph $\mathscr{G}^* = (V^*, \mathscr{E}_3^* \cup \cdots \cup \mathscr{E}_k^*)$ of $\mathscr{G}$, such that, simultaneously for $i \in \{3, \ldots, k\}$, $j \in \{2, \ldots, k-1\}$ and $u \in \{r+1, \ldots, jr\}$, we have

$$\begin{aligned} |V^*| &\geqslant c' \cdot m^{\frac{kr}{2(k-1)} + \delta}, \\ |\mathscr{E}_i^*| &\leqslant F_i \cdot m^{\frac{ir}{2(k-1)} + \delta \cdot i}, \\ p_{j,u}(V^*) &\leqslant F' \cdot m^{u - \frac{(k-2)j}{2(k-1)} r + \delta \cdot j}, \\ s_{j,u}(\mathscr{G}_k^*) &\leqslant F \cdot m^{\frac{(k-2)j+2k}{2(k-1)} \cdot r - u + \delta \cdot (2k-j)}, \end{aligned} \tag{3.9}$$

for some appropriate positive constants c', F_i, F', F depending only on k and r. We show the following properties.

Claim 1. *Let* $0 < \delta < \frac{1}{4k(k-1)}$. *The following hold:*

(a) $|\mathscr{E}_i^*| = o(|V^*|)$, *for* $i = 3, \ldots, k-1$,

(b) $\min\{p_{j,u}(V^*), s_{j,u}(\mathscr{G}_k^*)\} = o(|V^*|)$, *for* $j = 2, \ldots, k-1$ *and* $u = r+1, \ldots, jr$.

Proof. For the proof of (a), we observe that

$$\frac{ir}{2(k-1)} + \delta \cdot i < \frac{ir}{2(k-1)} + \frac{1}{4(k-1)} < \frac{kr}{2(k-1)} + \delta;$$

hence $|\mathscr{E}_i^*| = o(|V^*|)$. For the proof of (b), assume the contrary, *i.e.*, that there exist $j \in \{2, \ldots, k-1\}$ and $u \in \{r+1, \ldots, jr\}$ such that the following two inequalities hold:

$$\begin{aligned} u - \frac{(k-2)jr}{2(k-1)} + \delta \cdot j &\geqslant \frac{kr}{2(k-1)} + \delta, \\ \frac{(k-2)j+2k}{2(k-1)} \cdot r - u + \delta \cdot (2k-j) &\geqslant \frac{kr}{2(k-1)} + \delta. \end{aligned}$$

Let us abbreviate $T := \frac{(k-2)j+k}{2(k-1)} \cdot r$. From the two inequalities, it follows that

$$T - \delta \cdot (j-1) \leqslant u \leqslant T + \delta \cdot (2k-j-1).$$

We now observe that $T = \left(j - \frac{k(j-1)}{2(k-1)}\right) \cdot r$ (for $j = 2, \ldots, k-1$) is not an integer. The reason is that $\gcd(k, k-1) = 1$, $\gcd(k-1, r) = 1$ by assumption and $j - 1 < k - 1$. Furthermore, T is at least $\frac{1}{2(k-1)}$ away from the nearest integer.

On the other hand, we have by the choice of δ that $\delta \cdot (j-1) < \frac{1}{2(k-1)}$ as well as $\delta \cdot (2k-j-1) < \frac{1}{2(k-1)}$; hence the interval corresponding to the lower and upper bound on u contains no integer. Since u is an integer, we have obtained a contradiction.

Given the hypergraph $\mathscr{G}^*$, we obtain a k-uniform induced subhypergraph of $\mathscr{G}^*$ without 2-cycles as follows. First, we remove all i-element edges for $i < k$ by removing one vertex

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

from each such edge. We then apply the following procedure for $j = 2, \ldots, k-1$ and $u = r+1, \ldots, jr$.

If $p_{j,u}(V^*) \leqslant s_{j,u}(\mathcal{G}_k^*)$, then we remove for every j-element subset $J \subseteq V^*$ with $p(J) = u$ one vertex $v \in J$.

Otherwise, we remove for every $(2,j)$-cycle $\{E_1, E_2\}$ in $\mathcal{G}^*$ with $p(E_1 \cap E_2) = u$ one vertex of the cycle.

With respect to Section 4 we note that, given the hypergraph $\mathcal{G}^*$, the above procedure can be carried out in polynomial time. We obtain an induced subhypergraph $\mathcal{G}^{**}$ of $\mathcal{G}^*$. By construction, $\mathcal{G}^{**}$ is k-uniform. Claim 1 shows that $\mathcal{G}^{**}$ has at least $\Omega(|V^*|)$ many vertices which is $\Omega\big(m^{\frac{kr}{2(k-1)}+\delta}\big)$. Furthermore, $\mathcal{G}^{**}$ has no 2-cycles. To see this, consider a $(2,j)$-cycle $\{E_1, E_2\}$ such that $E_1 \cap E_2 = J$ and $p(J) = u$. During the above procedure, we either delete one vertex from J or we delete one vertex of the cycle, depending on the values of $p_{j,u}(V^*)$ and $s_{j,u}(\mathcal{G}_k^*)$.

The number of edges in $\mathcal{G}^{**}$ is bounded from above by $O\big(m^{\frac{kr}{2(k-1)}+\delta \cdot k}\big)$. Thus, the average degree of the vertices in $\mathcal{G}^{**}$ is bounded by $O(m^{\delta \cdot (k-1)})$; hence the parameter t of $\mathcal{G}^{**}$ is $O(m^{\delta})$.

We apply Theorem 3.1 and obtain an independent set in $\mathcal{G}^{**}$ and thus in $\mathcal{G}$ of size at least

$$\Omega\left(\frac{m^{\frac{kr}{2(k-1)}+\delta}}{m^{\delta}} \cdot (\ln m^{\delta})^{\frac{1}{k-1}}\right) = \Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\right),$$

since $\delta > 0$ is a constant depending only on k. The vertices of this independent set yield the desired 0–1 matrix with k-wise independent columns. □

4. An algorithm

The result we have obtained so far is only an existence result. However, this existence proof can be derandomized. For the derandomization, we apply the following result from [6], which extends former investigations of Fundia [9] and which is the algorithmic analogue of Theorem 3.1.

Theorem 4.1 ([6]). *Let $k \geqslant 3$ be a fixed integer. Let $\mathcal{G} = (V, \mathcal{E})$ be a k-uniform hypergraph on n vertices with average degree at most t^{k-1}. If $\mathcal{G}$ has no 2-cycles, then one can find for any fixed $\delta > 0$ in time $O(|V| + |\mathcal{E}| + n^3/t^{3-\delta})$ an independent set of size at least $c(k,\delta) \cdot n/t \cdot (\ln t)^{1/(k-1)}$.*

Using this result and the method of conditional probabilities (see [5]), we can prove the following.

Theorem 4.2. *Let k and r be fixed integers such that $r \geqslant 2$, $k \geqslant 4$ is even, and $\gcd(k-1, r) = 1$. For every integer $m \geqslant 1$ one can find in polynomial time a 0–1 matrix with m rows and $\Omega\big(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\big)$ columns such that each column contains exactly r ones and each set of k columns is linearly independent over $\mathbb{Z}_2$.*

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

Proof. The only place in the proof of Theorem 3.2 where we applied probabilistic arguments was in the existence proof of the hypergraph $\mathscr{G}^*$. Hence, it is enough to show that we can find such a hypergraph by an algorithm in polynomial time.

Let $C_m^r = \{a_1, \ldots, a_M\}$ (where $M = \binom{m}{r}$) be the set of all 0–1 column vectors of length m with exactly r ones. Associate with each column vector a_i a weight $p_i \in [0,1]$. For $i = 3, \ldots, k$, let $\mathscr{E}_i$ be the set of all i-element subsets $\{a_1, \ldots, a_i\} \in [C_m^r]^i$ such that $a_1 \oplus \cdots \oplus a_i = 0$. Moreover, for $j = 2, \ldots, k-1$ and $u = r+1, \ldots, jr$, let $P(j;u)$ be the set of all j-element subsets $J \in [C_m^r]^j$ with $p(J) = u$. Let $T(j;u)$ be the set of all pairs $\{E_1, E_2\} \in [\mathscr{E}_k]^2$, with $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$, where $j = 2, \ldots, k-1$ and $u = r+1, \ldots, jr$. We define the following potential function for $p := m^{-\frac{(k-2)r}{2(k-1)}+\delta}$ and some $0 < \delta < 1/(4k(k-1))$:

$$\begin{aligned} F(p_1, \ldots, p_M) \quad := \quad & 3^{pM/3} \cdot \prod_{i=1}^{M} \left(1 - \frac{2}{3} \cdot p_i\right) + \sum_{i=3}^{k} \frac{\sum_{E \in \mathscr{E}_i} \prod_{a_j \in E} p_j}{6k \cdot c_i \cdot p^i \cdot m^{ir/2}} \\ & + \sum_{j=2}^{k-1} \sum_{u=r+1}^{jr} \frac{\sum_{J \in P(j;u)} \prod_{a_i \in J} p_i}{6k^2 \cdot r \cdot C_2' \cdot p^j \cdot m^u} \\ & + \sum_{j=2}^{k-1} \sum_{u=r+1}^{jr} \frac{\sum_{\{E_1,E_2\} \in T(j;u)} \prod_{a_i \in E_1 \cup E_2} p_i}{6k^2 \cdot r \cdot C_1' \cdot p^{2k-j} \cdot m^{kr-u}}. \end{aligned}$$

Notice that in the three sum terms the denominators contain upper bounds on the sizes of the objects we want to control: *cf.* (3.6), (3.7) and (3.8).

We start by choosing $p_1 = \ldots = p_M = p$. We claim that, if m is large enough, then $F(p, \ldots, p) < 1$. Indeed, using $1 - x \leqslant e^{-x}$, we infer

$$\begin{aligned} F(p, \ldots, p) \quad \leqslant \quad & 3^{pM/3} \cdot \left(1 - \frac{2}{3} \cdot p\right)^M + \sum_{i=3}^{k} \frac{1}{6k} \\ & + \sum_{j=2}^{k-1} \sum_{u=r+1}^{jr} \frac{1}{6k^2 r} + \sum_{j=2}^{k-1} \sum_{u=r+1}^{jr} \frac{1}{6k^2 r} \\ \leqslant \quad & \left(\frac{3}{e^2}\right)^{pM/3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ < \quad & 1. \end{aligned}$$

In each step $i = 1, \ldots, M$, the algorithm chooses $p_i = 0$ or $p_i = 1$ to minimize the value of the current potential. Finally, we obtain $p_1, \ldots, p_M \in \{0,1\}$. As the potential $F(p_1, \ldots, p_M)$ is linear in each p_i, we have $F(p_1, \ldots, p_M) \leqslant F(p, \ldots, p) < 1$. Let $V^* = \{a_i \in C_m^r \mid p_i = 1\}$ be the set of the chosen column vectors. By definition of the potential F, we have $|V^*| \geqslant pM/3 = \Omega\left(m^{\frac{kr}{2(k-1)}+\delta}\right)$. For $i = 3, \ldots, k$, the number of i-element edges in the induced subhypergraph on V^* is at most

$$6k \cdot c_i \cdot m^{\frac{ir}{2}} \cdot p^i = o\left(m^{\frac{kr}{2(k-1)}+\delta}\right),$$

as otherwise $F(p_1, \ldots, p_M) > 1$. Similarly, using Claim 1(b), we can verify that, by the choice of $\delta > 0$, for $j = 2, \ldots, k-1$ and $u = r+1, \ldots, jr$, either the number of all j-element

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

subsets $J \in [V^*]^j$ with $p(J) = u$ is at most

$$6k^2 \cdot r \cdot C_2' \cdot m^u \cdot p^j = o\left(m^{\frac{kr}{2(k-1)}+\delta}\right)$$

or the number of all pairs $\{E_1, E_2\}$ with $E_1, E_2 \in [V^*]^k \cap \mathscr{E}_k$, with $|E_1 \cap E_2| = j$ and $p(E_1 \cap E_2) = u$ is at most

$$6k^2 \cdot r \cdot C_1' \cdot p^{2k-j} \cdot m^{kr-u} = o\left(m^{\frac{kr}{2(k-1)}+\delta}\right).$$

Hence, our algorithm finds an appropriate hypergraph $\mathscr{G}^*$ and we can proceed as in the proof of Theorem 3.2, that is, deleting all i-element dependencies, $3 \leqslant i \leqslant k-1$, and all 2-cycles. We obtain a subset $V^{**} \subseteq V^*$ of column vectors with $|V^{**}| = \Omega\left(m^{\frac{kr}{2(k-1)}+\delta}\right)$, and the corresponding induced subhypergraph is k-uniform, contains at most $O\left(m^{\frac{kr}{2(k-1)}+k\delta}\right)$ edges and has no 2-cycles. By applying the algorithm from Theorem 4.1 we obtain in polynomial time an independent set of size at least

$$\Omega\left(m^{\frac{kr}{2(k-1)}} \cdot (\ln m)^{\frac{1}{k-1}}\right).$$

The vertices of this independent set yield the columns of the desired matrix. The running time is mainly given by the number of groups we have to control: dominating here are the at most $O(m^{(k-1)r})$ many $(2,2)$-cycles. □

In Theorem 4.2 the case for even values for k is given. The corresponding result for odd k can be obtained analogously. Indeed, for an arbitrary subset S of the row indices which has size $\lceil m/2 \rceil$, consider the set X of all column vectors with exactly r ones which have an odd number of ones in S. Then, $|X| \geqslant c \cdot \binom{m}{r}$ for some positive constant c, and no odd number of these column vectors adds to zero. Instead of starting the algorithm with the hypergraph on the vertex set $V = C_m^r$, we just start the algorithm on the induced subhypergraph on X. In the corresponding considerations only constant factors change.

Final remarks

1. It was pointed out by one of the referees that to prove Theorem 4.2 one can also proceed as follows. In our arguments in Section 3 for proving Theorem 3.2, all calculations remain valid if we pick the columns at random according to a $(2k-2)$-wise independent distribution. Namely, k-wise independence suffices for (3.6), $(k-1)$-wise independence suffices for (3.7) and $(2k-2)$-wise independence suffices for (3.8). Hence, with respect to Chebyshev's inequality for guaranteeing all four inequalities (3.9) with positive probability, $(2k-2)$-wise independent random variables suffice.

From [11] (see also [3]) we know that we can construct a k-wise independent probability space $\Omega \subseteq \{0,1\}^n$ of size $|\Omega| = O(n^{2k})$. Hence, using a $(2k-2)$-wise independent probability space $\Omega \subseteq \{0,1\}^{\binom{m}{r}}$ of size $|\Omega| = O(m^{r(4k-4)})$, we can find, by trying all $|\Omega|$ possibilities, a subset $V^* \subseteq V$ such that all four inequalities (3.9) simultaneously hold. Then we can proceed with the algorithm as in the proof of Theorem 3.2. This also yields a polynomial time algorithm for constructing a desired matrix.

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press

2. It might be interesting to investigate whether analogues of the results of Lubotzky, Phillips and Sarnak [16] and Margulis [17] for hypergraphs or other algebraic constructions (*cf.* the case $k = 4$ and $r = 2$) can be obtained in order to improve the known bounds on $N(m, k, r)$. However, no general simple explicit constructions are known for achieving the lower bound (1.1) (see [14]).

References

[1] Ajtai, M., Komlós, J., Pintz, J., Spencer, J. and Szemerédi, E. (1982) Extremal uncrowded hypergraphs. *J. Combin. Theory Ser. A* **32** 321–335.

[2] Ajtai, M., Komlós, J. and Szemerédi, E. (1981) A dense infinite Sidon sequence. *European J. Combin.* **2** 2–11.

[3] Alon, N., Babai, L. and Itai, A. (1986) A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms* **7** 567–583.

[4] Alon, N., Lefmann, H. and Rödl, V. (1991) On an anti-Ramsey type result. In *Sets, Graphs and Numbers*, Vol. 60 of *Colloquia Mathematica Societatis János Bolyai*, pp. 9–22.

[5] Alon, N. and Spencer, J. (1992) *The Probabilistic Method*, Wiley, New York.

[6] Bertram-Kretzberg, C. and Lefmann, H. (1997) The algorithmic aspects of uncrowded hypergraphs. In *Proc. 8th ACM–SIAM Symposium on Discrete Algorithms SODA*, pp. 296–304. Also to appear in *SIAM J. Comput.*

[7] Duke, R. A., Lefmann, H. and Rödl, V. (1995) On uncrowded hypergraphs. *Random Struct. Algorithms* **6** 209–212.

[8] Frankl, P. and Füredi, Z. (1986) Union-free families of sets and equations over fields. *J. Number Theory* **23** 210–218.

[9] Fundia, A. (1996) Derandomizing Chebychev's inequality to find independent sets in uncrowded hypergraphs. *Random Struct. Algorithms* **8** 131–147.

[10] Füredi, Z. (1991) Turán type problems. In *Surveys in Combinatorics*, Vol. 166 of *London Math. Soc. Lecture Notes*, Cambridge University Press, pp. 253–300.

[11] Karloff, H. and Mansour, Y. (1994) On construction of k-wise independent random variables. In *Proc. 26th Annual ACM Symposium on Theory of Computing* (*STOC*), pp. 564–573.

[12] Kohayakawa, Y., Kreuter, B. and Steger, A. (1998) An extremal problem for random graphs and the number of graphs with large even-girth. *Combinatorica* **18** 101–120.

[13] Lazebnik, F., Ustimenko, V. A. and Woldar, A. J. (1995) A new series of dense graphs of high girth. *Bull.* (*New Series*) *Amer. Math. Soc.* **32** 73–79.

[14] Lefmann, H., Pudlák, P. and Savický, P. (1997) On sparse parity check matrices. *Designs, Codes and Cryptography* **12** 107–130.

[15] Lidl, R. and Niederreiter, H. (1994) *Introduction to Finite Fields and their Applications*, Cambridge University Press.

[16] Lubotzky, A., Phillips, R. and Sarnak, P. (1988) Ramanujan graphs. *Combinatorica* **8** 261–277.

[17] Margulis, G. A. (1988) Explicit group theoretical construction of combinatorial schemes and their application to the design of expanders and concentrators. *J. Probl. Inform. Transmission* **24** 39–46.

[18] Pudlák, P. and Rödl, V. (1993) Modified ranks of tensors and the size of circuits. In *Proc. 25th ACM Symposium on the Theory of Computing* (*STOC*), pp. 523–531.

[19] Sipser, M. and Spielman, D. A. (1994) Expander codes. In *Proc. 35th IEEE Symposium on Foundations of Computer Science* (*FOCS*), pp. 566–576.

[20] Rödl, V. and Siňajová, E. (1994) Note on independent sets in Steiner systems. *Random Struct. Algorithms* **5** 183–190.

[21] Spielman, D. A. (1995) Linear-time encodable and decodable error-correcting codes. In *Proc. 27th ACM Symposium on the Theory of Computing* (*STOC*), pp. 388–397.

[22] Spielman, D. A. (1995) Computationally efficient error-correcting codes and holographic proofs. PhD thesis, MIT.

https://doi.org/10.1017/S0963548399004058 Published online by Cambridge University Press