S-constrained random matrices

Sylvain Gravier[1†] and Bernard Ycart[2‡]

[1]*CNRS - UJF, ERTé "Maths à Modeler", Groupe de Recherche GéoD - Laboratoire Leibniz, 46, avenue Félix Viallet, 38031 Grenoble Cedex (France)*
[2]*LMC, CNRS UMR 5523, BP 53, 38041 Grenoble cedex 9, FRANCE*

Let S be a set of d-dimensional row vectors with entries in a q-ary alphabet. A matrix M with entries in the same q-ary alphabet is S-constrained if every set of d columns of M contains as a submatrix a copy of the vectors in S, up to permutation. For a given set S of d-dimensional vectors, we compute the asymptotic probability for a random matrix M to be S-constrained, as the numbers of rows and columns both tend to infinity. If n is the number of columns and $m = m_n$ the number of rows, then the threshold is at $m_n = \alpha_d \log(n)$, where α_d only depends on the dimension d of vectors and not on the particular set S. Applications to superimposed codes, shattering classes of functions, and Sidon families of sets are proposed. For $d = 2$, an explicit construction of a S-constrained matrix is given.

Keywords: random matrix, Poisson approximation, superimposed code, shattering, VC dimensions, Sidon families

1 Introduction

We propose S-constrained matrices as a unifying framework for two seemingly remote notions, one in the field of cryptography (superimposed codes), the other one in information theory (shattering classes of functions).

Definition 1 *Let q, d and s be three fixed integers. Let $S = \{\eta_1, \dots, \eta_s\}$ be a set of d-dimensional row vectors, with entries in the q-ary alphabet $\{0, \dots, q-1\}$.*

Let $m \geq d$ and $n \geq s$ be two integers. Let $M = (M_{i,j})$ be a $m \times n$ matrix with entries in the same q-ary alphabet.

The matrix M is said to be S-constrained if for every subset $\mathbf{j}$ of $[n] = \{1, \dots, n\}$, with $|\mathbf{j}| = d$ elements, there exist s indices $i_1, \dots, i_s$ in $[m]$ such that for all $h = 1, \dots, s$ and for all $j \in \mathbf{j}$,

$$M_{i_h,j} = \eta_h(j) \ .$$

In other words, every subset of d columns of the matrix M must contain as a submatrix a copy of all row vectors in S, up to permutation.

The matrices considered here can be interpreted either as q-ary codes (see Cohen and Schaathun's review (5)) or as classes of functions from $[n]$ into $\{0, \dots, q-1\}$ (see for instance Anthony and Bartlett (2)). In the first interpretation, the *columns* of the matrix are the words of the code, in the second one, the *rows* are the functions of the class.

To motivate Definition 1, recall ((16) p. 276) that a binary code is called (w, r)-*superimposed* if for every subsets of words W, R (sets of columns of the matrix) with respective cardinalities w, r, there exists a position (row index) on which every word of W is 1 and every word of R is 0. Let S be the set of all binary vectors with w ones and r zeros. The code is (w, r)-superimposed if and only if the corresponding matrix is S-constrained.

Consider now the interpretation of a q-ary matrix as a class of functions (rows of the matrix). A set $\mathbf{j}$ of d coordinates (column indices), is *shattered* by the class, if the restrictions of the functions to those coordinates contain all possible q^d functions. The *Vapnik-Chervonenkis dimension* of the class is the size of the largest shattered set, its *testing dimension* is the maximal d such that all sets of size d are shattered (see (3)). Let S be the set of all q^d vectors of size d. The testing dimension of the class is $\geq d$ if and only if the corresponding matrix is S-constrained.

†Email: `Sylvain.Gravier@imag.fr`
‡Email: `Bernard.Ycart@imag.fr`

The aim of this paper is to obtain asymptotic bounds on the size of S-constrained matrices, as the number of columns n tends to infinity, and the number of rows $m = m_n$ increases as a function of n. The problem of finding bounds on matrices satisfying certain constraints, has given rise to an extensive litterature: see *e.g.* Kim *et al.* (16) for (w, r)-superimposed codes, Cohen and Schaathun (5; 6) for separating codes, Li *et al.* (17) for hashing codes. The dual problem of finding matrices *not containing* any copy of S has also been considered by a number of authors, following Sauer (20): see Steele (21) for matrices not shattering any sets of size d, and more recently Anstee *et al.* (1) on matrices with forbidden configurations.

We are mainly concerned with probabilistic bounds, derived from examining the threshold of the desired property for random matrices: see (6; 16; 17) for comparisons between probabilistic and deterministic bounds. See also (22; 23), for probabilistic bounds on classes of binary functions under different random models.

Our main result gives the asymptotic probability for a random matrix to be S-constrained. By *random matrix*, we mean a matrix whose entries are mutually independent and uniformly distributed on the alphabet $\{0, \ldots, q-1\}$. Its distribution is denoted by $\mathbb{P}$.

Theorem 1 *Let S be a set of s q-ary vectors of size d. Denote by α_d the following (positive) real.*

$$\alpha_d = -\frac{d}{\log(1 - q^{-d})} \ . \tag{1}$$

Assume $m = m_n$ is such that:

$$\lim_{n\to\infty} m_n - \alpha_d \log n = c \ , \tag{2}$$

where c is a real constant. Let M be q-ary random matrix with m_n rows and n columns. Then:

$$\lim_{n\to\infty} \mathbb{P}(M \textit{ is } S\textit{-constrained}\,) = \exp\left(-\frac{s}{d!}(1 - q^{-d})^c\right) \ . \tag{3}$$

As a consequence of Theorem 1, $m_n = \alpha_d \log n$ is a probabilistic bound for S-constrained matrices: if the number of rows is such that that $m_n - \alpha_d \log n$ tends to $-\infty$, then with high probability (*w.h.p.*), a random matrix is not S-constrained. If $m_n - \alpha_d \log n$ tends to $+\infty$, it is S-constrained *w.h.p.* Observe that the threshold $\alpha_d \log n$ only depends on the dimension of the vectors in S, and not on any particular set S. Thus if $m_n - \alpha_d \log n$ tends to $+\infty$ a random matrix, interpreted as a code, is (w, r)-superimposed for any w, r such that $w + r = d$. Interpreted as a class of functions, it has testing dimension at least d. We shall see in Section 3 another application to Sidon families of sets.

The question naturally arises of whether $m_n = \alpha_d \log n$ is optimal. We will answer it by the negative in section 4. In the case $d = 2$, we construct a S-constrained matrix having $m_n \leq \beta_2 \log n$ rows, with

$$\beta_2 = \frac{q^2}{\log 2} < \alpha_2 = -\frac{2}{\log(1 - q^{-2})} \ .$$

The problem of finding bounds for the size of S-constrained matrices has been investigated in the different contexts of combinatorics, cryptography and information theory. Deterministic and probabilistic bounds for q-ary superimposed codes are given by (16). D'yachkov and Rykov (8) used a probabilistic approach to construct superimposed codes in the binary case. Cohen and Schaathun (5) propose a very thorough review on bounds for separating codes, completed by (6). Other results concerning Sidon families, cover-free families and superimposed codes can be found in (10; 8; 13; 19; 12; 9; 16).

The rest of the paper is organised as follows. Section 2 is dedicated to the proof of Theorem 1, based on a classical Poisson approximation technique. Section 3 details the application of S-constrained binary matrices to Sidon families of sets. In Section 4, we construct an explicit S-constrained matrix in the case $d = 2$.

2 Proof of Theorem 1

Before proving Theorem 1, we will comment a few of its consequences.

The convergence in (3) expresses a rather sharp concentration result. Consider a random matrix with $m_n = \alpha \log n$ rows, where $\alpha \neq \alpha_d$ for all d, and let d be the largest integer such that $\alpha > \alpha_d$. Then if S is any set of d-dimensional vectors, the matrix will be S-constrained *w.h.p.* But for any set of $(d+1)$-dimensional vectors, it will not be S-constrained *w.h.p.*

For $m_n = \alpha_d \log n + c$, the probability to be S-constrained converges to a non trivial value, that does depend on the set S, but only as a decreasing function of its cardinality s: the more constraints are imposed, the less likely it is to satisfy them all. Table 1 below gives numerical values of the asymptotic probability (3), for $q = 2$, $c = 5$, $d = 1, \ldots, 10$, and for the two extreme values of $s = 1$ and $s = 2^d$. It turns out that there is relatively little difference between satisfying all 2^d possible constraints, or just a single one.

d	1	2	3	4	5	6	7	8	9	10
$s = 1$	0.9692	0.8881	0.9181	0.9703	0.9929	0.9987	0.9998	1.0000	1.0000	1.0000
$s = 2^d$	0.9394	0.6221	0.5047	0.6171	0.7965	0.9211	0.9759	0.9938	0.9986	0.9997

Tab. 1: Asymptotic probability for a random matrix to be S-constrained for a set of d-dimensional vectors, of size $s = 1$ or $s = 2^d$, for $q = 2$, $c = 5$ and $d = 1, \ldots, 10$.

Theorem 1 provides an existence result for S-constrained matrices: if $m_n = \alpha_d \log n + c$, then for n large enough, the probability for a random matrix to be S-constrained is strictly positive, hence there must exist S-constrained matrices of size $m_n \times n$. But it also provides an algorithmic way to construct such a matrix. Assume the parameters are such that the probability for a random matrix to be S-constrained is $1/2$, then drawing random matrices until one which is S-constrained is found, will output the desired matrix after 2 random drawings on average.

Theorem 1 is a Poisson approximation result (see Barbour et al. (4) for a general reference). The technique of proof, based on the Stein-Chen method, is quite standard: we shall use the results stated in Janson (14).

Proof: Let $\mathbf{j} \subset [n]$ be a set of d column indices. For any d-dimensional q-ary row vector η, we denote by $C_{\mathbf{j},\eta}$ the set of all $m \times n$ q-ary matrices such that there exists a row of M whose restriction to $\mathbf{j}$ is η. We denote by $C_{\mathbf{j}}$ the intersection of the $C_{\mathbf{j},\eta}$'s over all vectors $\eta \in S$

$$C_{\mathbf{j}} = \bigcap_{\eta \in S} C_{\mathbf{j},\eta} \;.$$

The set $C_{\mathbf{j}}$ is made of those matrices whose columns indexed by $\mathbf{j}$ contain a copy of all vectors in S. The set of S-constrained matrices is the intersection over all possible subsets of d indices $\mathbf{j}$, of the events $C_{\mathbf{j}}$.

$$C = \bigcap_{|\mathbf{j}|=d} C_{\mathbf{j}} \;.$$

Our aim is to compute the probability of C under the uniform distribution on all $m \times n$ q-ary matrices, denoted by $\mathbb{P}$. We begin with the probability of $C_{\mathbf{j}}$, for $\mathbf{j} \subset [n]$ such that $|\mathbf{j}| = d$. We denote by $\overline{B}$ the complement of an event B. Recall that $C_{\mathbf{j}} = \cap C_{\mathbf{j},\eta}$, hence $\overline{C_{\mathbf{j}}} = \cup \overline{C_{\mathbf{j},\eta}}$, where the union extends over all elements η of S. If $h \leq s$ and $\eta_1, \ldots, \eta_h$ are distinct elements of S, then the probability that among m random rows none of them coincides with one of the η_i's on $\mathbf{j}$ is:

$$\mathbb{P}(\,\overline{C_{\mathbf{j},\eta_1}} \cap \ldots \cap \overline{C_{\mathbf{j},\eta_h}}\,) = (1 - hq^{-d})^m \;. \tag{4}$$

By the formula giving the probability of a union (formula (1.5) p. 99 of Feller (11)), one gets

$$\mathbb{P}(\,\overline{C_{\mathbf{j}}}\,) = \sum_{h=1}^{s} (-1)^{h-1} \binom{s}{h} (1 - hq^{-d})^m \;.$$

If $m = m_n$ tends to infinity, this sum tends to 0 and the first term dominates.

$$\mathbb{P}(\,\overline{C_{\mathbf{j}}}\,) = s(1 - q^{-d})^{m_n}(1 + o(1)) \;.$$

Recall that $\alpha_d = -d/\log(1 - q^{-d})$. For $m_n = \alpha_d \log n + c + o(1)$,

$$\mathbb{P}(\,\overline{C_{\mathbf{j}}}\,) = s(1 - q^{-d})^c \, n^{-d}(1 + o(1)) \;. \tag{5}$$

Let us now introduce the integer-valued random variable X which counts the number of failed events among the $C_{\mathbf{j}}$'s.

$$X = \sum_{|\mathbf{j}|=d} \mathbb{I}_{\overline{C_{\mathbf{j}}}} \;,$$

where $\mathbb{I}_B$ denotes the indicator function of an event B. Obviously, a random matrix is S-constrained, iff $X = 0$. The expectation of X is

$$\mathbb{E}(X) = \binom{n}{d} \mathbb{P}(\overline{C}_{\mathbf{j}}) \; . \tag{6}$$

As n tends to infinity,

$$\binom{n}{d} = \frac{n^d}{d!}(1 + o(1)) \; .$$

Therefore, for $m_n = \alpha_d \log n + c + o(1)$,

$$\lim_{n\to\infty} \mathbb{E}(X) = \frac{s}{d!}(1 - q^{-d})^c \; . \tag{7}$$

Comparing (7) with (3), there remains to prove that

$$\lim_{n\to\infty} \mathbb{P}(X = 0) = \lim_{n\to\infty} \exp(-\mathbb{E}(X)) \; ,$$

We will prove that a Poisson approximation holds for X. The family of indicators $(\mathbb{I}_{\overline{C_{\mathbf{j}}}})$ is *dissociated*, in the sense of Janson (14) p. 10: the two sets of random variables $\{\mathbb{I}_{\overline{C_{\mathbf{j}}}} , \; \mathbf{j} \in J_1\}$ and $\{\mathbb{I}_{\overline{C_{\mathbf{j}'}}} , \; \mathbf{j}' \in J_2\}$ are independent whenever every $\mathbf{j} \in J_1$ is disjoint from every $\mathbf{j}' \in J_2$. Denote by Γ the set of all $\mathbf{j} \subset [n]$ with $|\mathbf{j}| = d$. For $\mathbf{j} \in \Gamma$, denote by $\Gamma_{\mathbf{j}}$ the set of all $\mathbf{j}'$ such that $\mathbf{j} \cap \mathbf{j}' \neq \emptyset$. By Theorem 4 p. 10 of (14), the total variation distance between the distribution of X and the Poisson distribution with parameter $\mathbb{E}(X)$ is bounded above by

$$(1 \wedge \mathbb{E}(X)^{-1}) \left(\sum_{\mathbf{j}\in\Gamma} \sum_{\mathbf{j}'\in\Gamma_{\mathbf{j}}} \mathbb{P}(\overline{C_{\mathbf{j}}})\mathbb{P}(\overline{C_{\mathbf{j}'}}) + \sum_{\mathbf{j}\in\Gamma} \sum_{\mathbf{j}'\in\Gamma_{\mathbf{j}}\setminus\{\mathbf{j}\}} \mathbb{P}(\overline{C_{\mathbf{j}}} \cap \overline{C_{\mathbf{j}'}}) \right) \tag{8}$$

The result will follow by proving that each of the two sums in (8) converges to zero. The first sum has $O(n^{2d-1})$ terms, each of order $O(n^{-2d})$, by (5). We decompose the second sum according to the number of elements in $\mathbf{j} \cap \mathbf{j}'$ as follows:

$$\sum_{\mathbf{j}\in\Gamma} \sum_{\mathbf{j}'\in\Gamma_{\mathbf{j}}\setminus\{\mathbf{j}\}} \mathbb{P}(\overline{C_{\mathbf{j}}} \cap \overline{C_{\mathbf{j}'}}) = \sum_{h=1}^{d-1} \Delta_h \; ,$$

where

$$\Delta_h = \sum_{|\mathbf{j}\cap\mathbf{j}'|=h} \mathbb{P}(\overline{C}_{\mathbf{j}} \cap \overline{C}_{\mathbf{j}'}) \; .$$

Clearly, there are $O(n^{2d-h})$ terms in Δ_h. We have

$$\overline{C}_{\mathbf{j}} \cap \overline{C}_{\mathbf{j}'} = \left(\bigcup_{\eta\in S} \overline{C}_{\mathbf{j},\eta} \right) \bigcap \left(\bigcup_{\zeta\in S} \overline{C}_{\mathbf{j}',\zeta} \right) = \bigcup_{\eta,\zeta} (\overline{C}_{\mathbf{j},\eta} \cap \overline{C}_{\mathbf{j}',\zeta}) \; .$$

The probability of any of these intersections is:

$$\mathbf{P}(\overline{C}_{\mathbf{j},\eta} \cap \overline{C}_{\mathbf{j}',\zeta}) = \begin{cases} (1 - 2q^{-d} + q^{-2d+h})^m & \text{if } \eta \equiv \zeta \text{ on } \mathbf{j} \cap \mathbf{j}' \\ (1 - 2q^{-d})^m & \text{otherwise.} \end{cases}$$

Hence

$$\Delta_h \leq a \, n^{2d-h}(1 - 2q^{-d} + q^{-2d+h})^m$$

for some positive a, not depending on n and m. For $m = m_n = \alpha_d \log n + c + o(1)$, there exists a positive constant b such that

$$\Delta_h \leq b \, n^{2d-h+\alpha_d \log(1-2q^{-d}+q^{-2d+h})} \; .$$

That Δ_h tends to zero follows from having a negative exponent, *i.e.*,

$$2d - h + \alpha_d \log(1 - 2q^{-d} + q^{-2d+h}) < 0 \tag{9}$$

for $d \geq 2$ and $h = 1, \ldots, d-1$. Indeed, the left hand side of (9) vanishes both for $h = 0$ and $h = d$. As a function of h, its second derivative is positive on $[0, d]$ hence it is strictly convex. Therefore it is strictly negative for all $h = 1, \ldots, d-1$. Hence Δ_h tends to zero with increasing n, which concludes the proof. □

3 Sidon families

Let $\mathcal{R} = \{R_1, \ldots, R_n\}$ be a family of subsets of $[m]$. To each of them, one can associate a m-dimensional binary column vector, whose i-th entry is 1 if $i \in R_j$ and 0 else. This defines the *incidence matrix* of the family, denoted by $M = (M_{i,j})$.

$$M_{i,j} = \mathbb{I}_{R_j}(i) \ .$$

Superimposed codes translate into *cover-free families*: $\mathcal{R}$ is a (w, r)*-cover-free family*, if no intersection of w members of the family is covered by a union of r others (see (18; 9; 7; 16)). As already observed, the incidence matrix of a (w, r)-cover-free family is S-superimposed, where S is the set of vectors having w ones and r zeros. The notion we are studying in this section does not translate as straightforwardly.

Definition 2 *A set family $\mathcal{R}$ is a k-*Sidon family *if all the k-term unions are distinct :*

$$\bigcup_{h=1}^{k} R_{j_h} \neq \bigcup_{h=1}^{k} R_{j'_h} \ ,$$

whenever $\{j_1, \ldots, j_k\} \neq \{j'_1, \ldots, j'_k\}$.

The notion was defined in (15) where the authors used the terminology 'UD_k code'. One can also find in the literature the term 'k-superimposed family' in (8). We adopt here the terminology of (13).

Contrarily to cover-free families, there is no set of vectors S such that the family is k-Sidon if and only if its incidence matrix is S-constrained. Proposition 1 gives a necessary and a sufficient condition.

Proposition 1 *Let k be a positive integer. Let S_0 be the singleton containing only the null vector with k entries. Let S_1 be the set of all vectors of length $(k+1)$ having a single one and k zeros.*

1. *If a family of sets is k-Sidon, then in its incidence matrix, at most one set of k columns does not contain S_0.*

2. *If the incidence matrix is S_1-constrained, then the family of sets is k-Sidon.*

Proof: Consider first the necessary condition. A set of k columns does not contain the null vector, if and only if the union of the k sets is $[m]$. If the family is k-Sidon, this can happen at most once.

Let us now turn to the sufficient condition. Consider two sets of indices $j_1, \ldots, j_k$ and $j'_1, \ldots, j'_k$. With no loss of generality, assume that $j_1 \neq j'_1$. Since the incidence matrix is S_1-constrained, there exists an index $i \in [m]$ such that $M_{i,j_1} = 1$ and $M_{i,j'_1} = \ldots = M_{i,j'_k} = 0$: the element i belongs to R_{j_1}, but not to $R_{j'_1} \cup \ldots \cup R_{j'_k}$. Hence the family is k-Sidon. □

Consider a random binary matrix in the sense of Theorem 1, with m_n rows and n columns. It is the incidence matrix of a family of n sets, provided its columns are distinct. This holds *w.h.p.* if m_n is large compared to $(2/\log 2)\log n$. As a consequence of point *2* of Proposition 1 and Theorem 1, if $m_n - \alpha_{k+1}\log n$ tends to $+\infty$, then the family is k-Sidon *w.h.p.* From point *1* of Proposition 1, for a k-Sidon family, the variable X counting the number of failures of the property of having a copy of S_0 is either 0 or 1. It follows from the Poisson approximation of Theorem 1, that the threshold for '$X \leq 1$' is the same as for '$X = 0$', *i.e.* $\alpha_k \log n$: if $m_n - \alpha_k \log n$ tends to $-\infty$, then the family will not be k-Sidon *w.h.p.*

4 Explicit construction

Given a set $S = \{\eta_1, \ldots, \eta_s\}$ of s 2-dimensional vectors with entries in a q-ary alphabet, we will construct a $m \times n$ S-constrained matrix M with $m \leq \frac{q^2}{\log 2}\log n$. Notice that the size of this matrix is an improvement of the probabilistic bound obtained in Theorem 1. Indeed, the ratio

$$\frac{q^2/\log 2}{\alpha_2} = \frac{q^2/\log 2}{-2/\log(1-q^{-2})}$$

is smaller than 1 for all q and tends to $1/(2\log 2) \simeq 0.72$ as q increases.

Let k be a positive integer. We start with a pattern matrix P having $k+1$ rows and 2^k columns. Its first row is null. Its rows with indices $2, \ldots, k+1$ are formed by all 2^k binary column vectors, ranked in alphabetical order. Here is the matrix P for $k = 3$.

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

We denote by $\overline{P}$ the matrix $1 - P$.

Let S_1 be the set of all vectors (a, b) in S such that (b, a) also belongs to S, and $S_2 = S \setminus S_1$. Let s_1 and s_2 denote the respective cardinalities of S_1 and S_2. We will use as an example

$$S = \{(0,2),(2,0),(2,1)\}\ , \quad S_1 = \{(0,2),(2,0)\}\ , \quad S_2 = \{(2,1)\}\ .$$

Given a set X of two dimensional vectors, we denote by $A(X)$ (resp.: $B(X)$) the column vector of their first (resp.: second) coordinates. In our example:

$$A(S_1) = \binom{0}{2}\ ,\ B(S_1) = \binom{2}{0}\ ,\ A(S_2) = (2)\ ,\ B(S_2) = (1)\ .$$

Let U be the $(s_1 + s_2)(k+1) \times 2^{k+1}$ matrix obtained from the pattern matrix P by replacing all 0's (resp. all 1's) by the $(s1 + s2) \times 2$ block

$$\begin{pmatrix} A(S_1) & B(S_1) \\ A(S_2) & B(S_2) \end{pmatrix} \text{ (resp.: } \begin{pmatrix} B(S_1) & A(S_1) \\ B(S_2) & A(S_2) \end{pmatrix}) \ .$$

In our example:

$$U = \left(\begin{array}{cc|cc|cc|cc|cc|cc|cc|cc}
0&2&0&2&0&2&0&2&0&2&0&2&0&2&0&2\\
2&0&2&0&2&0&2&0&2&0&2&0&2&0&2&0\\
2&1&2&1&2&1&2&1&2&1&2&1&2&1&2&1\\
\hline
0&2&0&2&0&2&0&2&2&0&2&0&2&0&2&0\\
2&0&2&0&2&0&2&0&0&2&0&2&0&2&0&2\\
2&1&2&1&2&1&2&1&1&2&1&2&1&2&1&2\\
\hline
0&2&0&2&2&0&2&0&0&2&0&2&2&0&2&0\\
2&0&2&0&0&2&0&2&2&0&2&0&0&2&0&2\\
2&1&2&1&1&2&1&2&2&1&2&1&1&2&1&2\\
\hline
0&2&2&0&0&2&2&0&0&2&2&0&0&2&2&0\\
2&0&0&2&2&0&0&2&2&0&0&2&2&0&0&2\\
2&1&1&2&2&1&1&2&2&1&1&2&2&1&1&2
\end{array}\right)$$

Similarly, if S_2 is nonempty, we define V as the $s_2(k+1) \times 2^{k+1}$ matrix obtained by replacing in $\overline{P}$ all 0's (respectively 1's) by the $s_2 \times 2$ block:

$$(A(S_2), B(S_2)) \quad \text{(resp.: } (B(S_2), A(S_2)))\ .$$

In our example:

$$V = \left(\begin{array}{cc|cc|cc|cc|cc|cc|cc|cc}
1&2&1&2&1&2&1&2&1&2&1&2&1&2&1&2\\
\hline
1&2&1&2&1&2&1&2&2&1&2&1&2&1&2&1\\
\hline
1&2&1&2&2&1&2&1&1&2&1&2&2&1&2&1\\
\hline
1&2&2&1&1&2&2&1&1&2&2&1&1&2&2&1
\end{array}\right)$$

If S_2 is empty, let $M_S = U$. If S_2 is not empty, let M_S be the matrix $\binom{U}{V}$. The matrix M_S has $(s_1 + 2s_2)(k+1)$ rows and 2^{k+1} columns. In any case, $(s_1 + 2s_2) \le q^2$. Hence if $n = 2^{k+1}$ is the number of columns, the number of rows is no larger than $(q^2/\log 2)\log n$. Observe moreover that contrarily to α_2, the bound now depends on S: if S is a singleton, the matrix M_S only has $(2/\log 2)\log n$ rows.

Theorem 2 *The matrix M_S is S-constrained.*

Proof: We must prove that any two columns contain a copy of all row vectors of S. Let j_1 and j_2 be two column indices. We will prove that there exist $s_1 + s_2$ row indices $i_1, \ldots, i_{s_1}, i'_1, \ldots, i'_{s_2}$ such that the submatrix of M_S indexed by $\{i_1, \ldots, i_{s_1}, i'_1, \ldots, i'_{s_2}\} \times \{j_1, j_2\}$ is either

$$\begin{pmatrix} A(S_1) & B(S_1) \\ A(S_2) & B(S_2) \end{pmatrix} \text{ or } \begin{pmatrix} B(S_1) & A(S_1) \\ A(S_2) & B(S_2) \end{pmatrix}\ .$$

Since, up to permutation of rows, $(A(S_1), B(S_1))$ is the same as $(B(S_1), A(S_1))$, this will be enough to prove that M_S is S-constrained.

Assume first that j_1 and j_2 have distinct parities. Then the submatrix indexed by $\{1,\ldots,s_1\}\times\{j_1,j_2\}$ is either $(A(S_1),B(S_1))$ or $(A(S_1),B(S_1))$. If j_1 is odd then the submatrix indexed by $\{s_1+1,\ldots,s_1+s_2\}\times\{j_1,j_2\}$ is $(A(S_2),B(S_2))$. If j_1 is even, then the submatrix indexed by $\{s(k+1)+1,\ldots,s(k+1)+s_2\}\times\{j_1,j_2\}$ is $(A(S_2),B(S_2))$.

Assume now that j_1 and j_2 have the same parity. Let $j_1'=\lceil j_1/2\rceil$ and $j_2'=\lceil j_2/2\rceil$. By definition of P, there exists an index i, $1<i\leq k+1$, such that its coefficients of order (i,j_1') and (i,j_2') are distinct. So, the submatrix of M_S indexed by $\{(i-1)s+1,\ldots,(i-1)s+s_1\}\times\{j_1,j_2\}$ is either $(A(S_1),B(S_1))$ or $(B(S_1),A(S_1))$. Moreover, if the coefficient of order (i,j_1') of P is 0, then the submatrix of M_S indexed by $\{(i-1)s+s_1+1,\ldots,(i-1)s+s_1+s_2\}\times\{j_1,j_2\}$ is $(A(S_2),B(S_2))$. Otherwise, the coefficient of order $(i+k+1,j_1')$ of P is 0, and the submatrix of M_S indexed by $\{(k+1)s+(i-1)s_2+1,\ldots,(k+1)s+is_2\}\times\{j_1,j_2\}$ is $(A(S_2),B(S_2))$. □

5 Concluding remarks

The bounds deduced from Theorem 1 are probabilistic. Possibly, one could obtain a deterministic algorithm to construct such S-constrained matrices using a derandomization procedure which would lead to an algorithm similar to that of Hwang and Sós (13). Nevertheless, it would be even more interesting to get an 'explicit' construction for all d as we did in Section 4 for $d=2$. Indeed, such a construction gives more information on the structure of extremal matrices.

Hwang and Sós (13) used the concept of *part-intersecting family*. Given a set S of cardinality N, a t-part-intersecting family is a family $\mathcal{F}$ of subsets of S such that for all $A,B\in\mathcal{F}$

$$|A\cap B|<\frac{1}{t}\min_{A,B\in\mathcal{F}}\{|A|,|B|\}\ .$$

To understand how this problem relates with ours, consider the simpler constraint that every intersection $A\cap B$ has cardinality 2. For the incidence matrix, this means that any two columns must contain two copies of the row vector $(1,1)$. The technique that was used in the proof of Theorem 1 relied on the fact that the elements of S were distinct, and it cannot be directly adapted to counting copies of a given row vector. So a natural generalisation of our setting would be to consider matrices such that any set of d columns contains, up to permutation, a copy of some fixed matrix, which could have distinct rows or not.

Another interesting extension would be to consider matrices such that any set of d columns contains a copy of at least one set of vectors among different sets $S_1,\ldots,S_k$. This is related to separating codes, in a similar way as S-constrained matrices are related to superimposed codes.

References

[1] R. Anstee, B. Fleming, Z. Füredi, and A. Sali. Color critical hypergraphs and forbidden configurations. In *Proc. EuroComb2005*, pages 117–122. DMTCS, 2005.

[2] M. Anthony and P.L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[3] M. Anthony, G. Brightwell, D. Cohen, and J. Shawe-Taylor. On exact specification by examples. In *Proc. 5th Annual ACM Workshop on Comput. Learning Theory*, pages 311–318. ACM Press, New York, 1992.

[4] A.D. Barbour, L. Holst, and S. Janson. *Poisson approximation*. Clarendon Press, Oxford, 1992.

[5] G.D. Cohen and H.G. Schaathun. Asymptotic overview on separating codes. Technical Report 248, Dep. Informatics, Univ. Bergen, 2003.

[6] G.D. Cohen and H.G. Schaathun. Upper bounds on separating codes. *IEEE Trans. Inf. Theory*, 50(6):1291–1295, 2004.

[7] A. De Bonis and U. Vaccaro. Constructions of generalized superimposed codes with applications to group testing and conflict resolution in multiple access channels. *Theor. Comput. Sci.*, 306(1-3):223–243, 2003.

[8] A.G. D'yachkov and V.V. Rykov. Bounds on the length of disjunctive codes. *Prob. Info. Transmission*, 18:166–171, 1983.

[9] K. Engel. Interval packing and covering in the Boolean lattice. *Comb. Probab. Comput.*, 5(4):373–384, 1996.

[10] P. Erdős, P. Frankl, and Z. Füredi. Families of finite sets in which no set is covered by the union of two others. *J. Combinatorial Theory*, 33:158–166, 1982.

[11] W. Feller. *An introduction to probability theory and its applications*, volume I. Wiley, London, 3rd edition, 1968.

[12] Z. Füredi. On r-cover free families. *J. Combinatorial Theory*, 73(1):172–173, 1996.

[13] F.K. Hwang and V.T. Sós. Non-adaptive hypergeometric group testing. *Stud. Sci. Math. Hung.*, 22(1-4):257–263, 1987.

[14] S. Janson. Coupling and Poisson approximation. *Acta Appl. Math.*, 34:7–15, 1994.

[15] W.H. Kautz and R.C. Singleton. Nonrandom binary superimposed codes. *IEEE Trans. Inf. Theory*, 10:363–377, 1964.

[16] H.K. Kim, V. Lebedev, and D.Y. Oh. Some new results on superimposed codes. *J. Combinatorial Designs*, 13(4):276–285, 2005.

[17] P.C. Li, G.H.J. van Rees, and R. Wei. Constructions of 2-cover-free families and related separating hash families. Submitted, 2005.

[18] C.J. Mitchell and F.C. Piper. Key storage in secure networks. *Disc. Appl. Math.*, 21(3):215–228, 1988.

[19] M. Ruszinkó. On the upper bound of the size of the r-cover-free families. *J. Combinatorial Theory*, 66(2):302–310, 1994.

[20] N. Sauer. On the density of families of sets. *J. Combinatorial Theory (A)*, 13:145–147, 1972.

[21] J.M. Steele. Existence of submatrices with all possible columns. *J. Combinatorial Theory*, 24(1):84–88, 1978.

[22] B. Ycart and J. Ratsaby. The VC dimension of k-uniform random hypergraphs. Rand. Struct. Algo., to appear, 2006.

[23] B. Ycart and J. Ratsaby. VC-dimensions of random function classes. Submitted, 2006.