Lower Bounds

Given $d \in [N]$, let $t(d, N)$ denote the minimum t for which a d-disjunct matrix with t rows and N columns exists. We study the behavior of the function $t(d, N)$.

1 Lower bounds for large d

In the previous lecture note, an exercise showed that $t(N, N) = N$, so we can assume $d \in [N - 1]$.

Exercise 1. Show that $t(d, N) \geq \min\{3, N\}$, for all $d \in [N]$.

The following result was attributed to Bassalygo by Dýachkov and Rykov [4].

Proposition 1.1 (Bassalygo – 1975)**.** *The following bound holds*

$$t(d, N) \geq \min \left\{ \binom{d+2}{2}, N \right\}. \tag{1}$$

Proof. We induct on d. Exercise 1 proved the base case $d = 1$. Consider $d \geq 2$ and a d-disjunct matrix $\mathbf{M}$ with $t = t(d, N)$ rows and N columns. Let $N(w)$ denote the number of columns of $\mathbf{M}$ with weight w. (The weight of a column is the number of 1s in it.) A row $i \in [t]$ is said to be *private* for a column j if j is the only column in the matrix having a 1 on row i. If column $\mathbf{M}^j$ has weight at most d, then it must have at least one private element. The total number of private elements of all columns is at most t. Hence,

$$\sum_{w=1}^{d} N(w) \leq t.$$

Let $w_{\max}$ denote the maximum column weight of $\mathbf{M}$. If $w_{\max} \leq d$ then $N = \sum_w N(w) \leq t$. Now, suppose $w_{\max} \geq d + 1$ and consider a column $\mathbf{M}^j$ with weight equal to $w_{\max}$. If we remove column $\mathbf{M}^j$ and all rows i for which $m_{ij} = 1$, we are left with a $(d-1)$-disjunct matrix with $t - w_{\max}$ rows and $N - 1$ columns. Thus, $t - w_{\max} \geq t(d - 1, N - 1)$ which along with the induction hypothesis implies

$$t - (d + 1) \geq \min \left\{ \binom{d+1}{2}, N - 1 \right\}.$$

The proposition follows. □

Note that $t(d, N) \leq N$ is a trivial upper bound: the $N \times N$ identity matrix is d-disjunct. Bassalygo's bound says that if $\binom{d+2}{2} \geq N$ then we cannot do better than the identity matrix. Next, we consider the "small d" cases.

2 Lower bounds for small d

2.1 The $d = 1$ case

Consider a $t \times N$ binary matrix $\mathbf{M}$. Its columns can naturally be viewed as a family of subsets of $[t]$. The collection of columns of a 1-disjunct matrix satisfies the property that no set in the family is contained in another set in the family. Such a family is called an *anti-chain* in partially order set theory [2]. A classic (topology) lemma by Sperner in 1928 [5, 9] states that the maximum size of such an anti-chain is $\binom{t}{\lfloor t/2 \rfloor}$. Since the proof of Sperners lemma is short and illustrates a nice (probabilistic) technique, we reproduce it here.

Lemma 2.1 (Sperner lemma)**.** *Let $\mathcal{F}$ be a collection of subsets of $[t]$ such that no member of F is contained in another member of $\mathcal{F}$. Then, $|F| \leq \binom{t}{\lfloor t/2 \rfloor}$. Equality can be reached by picking $\mathcal{F} = \binom{[t]}{\lfloor t/2 \rfloor}$.*

Proof. Pick uniformly a random permutation π of $[t]$. For each member $F \in \mathcal{F}$, let A_F be the event that F is a prefix of π. For example, if $\pi = 3, 4, 1, 5, 2$ then $\{1, 3, 4\}$ is a prefix of π. Let $k = |F|$, then

$$\text{Prob}[A_F] = \frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}} \geq \frac{1}{\binom{[t]}{\lfloor t/2 \rfloor}}.$$

Because no member of $\mathcal{F}$ is contained in another, the events A_F are mutually disjoint. Thus,

$$1 \geq \sum_{F \in \mathcal{F}} \text{Prob}[A_F] \geq \frac{|\mathcal{F}|}{\binom{[t]}{\lfloor t/2 \rfloor}},$$

which completes the proof. $\square$

2.2 The Erdős-Frankl-Füredi technique

A subset $F \subseteq [t]$ is called a *private subset* of column $\mathbf{M}^j$ if $F \subseteq \mathbf{M}^j$ and $F \not\subseteq \mathbf{M}^{j'}$ for any $j' \neq j$. We first need the following lemma (Lemma 9.1 from Erdős-Frankl-Füredi [6]).

Lemma 2.2. *Let $\mathbf{M}$ be a $t \times N$ d-disjunct matrix. Fix a positive integer $w \leq t$. Let $\mathcal{C}$ denote the set of all columns of $\mathbf{M}$. Let C be any column in $\mathcal{C}$ which has no private w-subset. Consider any $k \geq 0$ other columns $C_1, \dots, C_k$ of $\mathbf{M}$. We have*

$$\left| C \setminus \bigcup_{j=1}^{k} C_j \right| \geq (d-k)w + 1. \tag{2}$$

In particular, if $\mathbf{M}$ has at least $d+1$ columns $C_1, \dots, C_{d+1}$ none of which have any private w-subset, then

$$\left| \bigcup_{j=1}^{d+1} C_j \right| \geq \frac{1}{2}(d+1)(dw+2). \tag{3}$$

Proof. If (2) does not hold, then C can be covered by the union of the $C_1, \ldots, C_k$ and $(d-w)$ other columns, contradicting the fact that $\mathbf{M}$ is d-disjunct. To prove (3), we apply (2) as follows.

$$\begin{aligned}
\left|\bigcup_{j=1}^{d+1} C_j\right| &= |C_1| + |C_2 \setminus C_1| + \cdots + |C_{d+1} \setminus C_1 \cup \cdots \cup C_d| \\
&\geq (dw+1) + ((d-1)w+1) + \cdots + (w+1) + 1 \\
&= \frac{d}{2}(d+1)w + (d+1) \\
&= \frac{1}{2}(d+1)(dw+2).
\end{aligned}$$

□

Theorem 2.3 (Füredi [7]). *For $N \geq d \geq 2$ and any d-disjunct matrix $\mathbf{M}$ with t rows and N columns, we have*

$$N \leq d + \binom{t}{\left\lceil \frac{t-d}{\binom{d+1}{2}} \right\rceil}.$$

Proof. Fix a non-negative integer $w \leq t/2$. Let $\mathcal{C}_w$ be the sub-collection of columns of $\mathbf{M}$ each of which has a private w-subset, and $\mathcal{C}_{<w}$ be the sub-collection of columns of $\mathbf{M}$ each of which has weight $< w$. Let $\mathcal{D}_w$ be a collection of private w-subsets of the sets in $\mathcal{C}_w$ where we just take one arbitrary private w-subset of each member of $\mathcal{C}_w$ to put in $\mathcal{D}$. Then, $\mathcal{D} \cup \mathcal{C}_{<w}$ forms an anti-chain, and the same technique used in the proof of Sperners lemma above can easily be used to show that, for any $|\mathcal{C}_w \cup C_{<w}| = |\mathcal{D} \cup C_{<w}| \leq \binom{t}{w}$. Now, if there were at least $d+1$ columns not in $C_w \cup C_{<w}$, then by Lemma 2.2 the union of columns **not** in $\mathcal{C}_w \cup \mathcal{C}_{<w}$ has cardinality at least $\frac{1}{2}(d+1)(dw+2)$. Suppose we are able to choose w such that $\frac{1}{2}(d+1)(dw+2) \geq t+1$ then we reach a contradiction, in which case we can conclude that $N \leq d + |C_w \cup C_{<w}| \leq d + \binom{t}{w}$. By setting $w = \left\lceil \frac{t+1-(d+1)}{\binom{d+1}{2}} \right\rceil$ we can be assured that $w \leq t/2$, and $\frac{1}{2}(d+1)(dw+2) \geq t+1$. □

Exercise 2. Show the missing piece in the above proof that, for any integer $1 \leq w \leq t/2$, $|\mathcal{C}_w \cup \mathcal{C}_{<w}| \leq \binom{t}{w}$.

Corollary 2.4 (Asymptotic lower bound for $t(d,N)$). *When $2 \leq d$ and $\binom{d+2}{2} < N$, we have*

$$t(d,N) \geq \frac{(d+1)^2}{12 \log d} \log N = \Omega\left(\frac{d^2}{\log d} \log N\right). \tag{4}$$

For $N \to \infty$ and $d \to \infty$, we have

$$t(d,N) \geq \frac{d^2}{4 \log d} \log N(1 + o(1)). \tag{5}$$

Proof. Note that $\frac{t}{2(t-d)} \leq 1$ because $t \geq \binom{d+2}{2}$. And, $\frac{t-d}{\binom{d+1}{2}} \leq \frac{2t}{d^2}$ is easy to verify, which implies $\left\lceil \frac{t-d}{\binom{d+1}{2}} \right\rceil \leq$

$\frac{2t}{d^2}$. We thus can bound

$$\begin{aligned}
\log \binom{t}{\left\lceil \frac{t-d}{\binom{d+1}{2}} \right\rceil} &\le \left\lceil \frac{t-d}{\binom{d+1}{2}} \right\rceil \log \left(\frac{te}{\left\lceil \frac{t-d}{\binom{d+1}{2}} \right\rceil} \right) \\
&\le \frac{2t}{d^2} \log \left((d+1)de \frac{t}{2(t-d)} \right) \\
&\le \frac{2t}{d^2} \log((d+1)^3) \\
&\le \frac{6t}{d^2} \log(d+1)
\end{aligned}$$

Secondly, when $\binom{d+2}{2} < N$ we have $\log(N-d) \ge \frac{1}{2}\log N$. Consequently,

$$\frac{1}{2}\log N \le \log(N-d) \le \frac{6t}{d^2}\log(d+1),$$

and (4) follows. The relation (5) is straightforward to verify. □

2.3 The Ruszinkó technique

Ruszinkó [8] devised a relatively simpler argument for proving a lowerbound for $t(d,N)$. The argument was slightly simplified in Alon-Asodi [1] although the bound shown in Alon-Asodi is slight worse. We present the simpler argument here.

As long as there is still a column in $\mathbf{M}$ with weight $\ge 2t/d$, remove the column along with all rows in which the column has 1s. When the process is finished, there were at most $d/2$ columns removed and all the remaining columns have weight $< 2t/d$. Each of the remaining columns must have a private $\lceil 4t/d^2 \rceil$-subset. Hence, the number of remaining columns is at most the number of subsets of $[t]$ of size $\lceil 4t/d^2 \rceil$. In other words,

$$N - d/2 \le \binom{t}{\lceil 4t/d^2 \rceil}.$$

When $t \ge \binom{d+2}{2}$, we have $t \ge d^2/2$. Thus, $\lceil 4t/d^2 \rceil \le 4t/d^2 + 1 \le 4t/d^2 + 2t/d^2 = 6t/d^2$. Furthermore, $N - d/2 \ge \sqrt{N}$. Consequently,

$$\frac{1}{2}\log N \le \log(N-d/2) \le \frac{6t}{d^2}\log\left(\frac{te}{4t/d^2}\right) < \frac{12t}{d^2}\log d.$$

We conclude that $t \ge \frac{d^2}{24\log d}\log N$.

2.4 The D'yachkov-Rykov technique

This technique [3] yields the best asymptotic bound of $t \ge \frac{d^2}{2\log d}\log N(1+o(1))$, but it is analytically complicated and the result requires $N \to \infty$ and $d \to \infty$ to work.

References

[1] N. Alon and V. Asodi, *Learning a hidden subgraph*, SIAM J. Discrete Math., 18 (2005), pp. 697–712 (electronic).

[2] B. Bollobás, *Combinatorics*, Cambridge University Press, Cambridge, 1986. Set systems, hypergraphs, families of vectors and combinatorial probability.

[3] A. G. D′yachkov and V. V. Rykov, *Bounds on the length of disjunctive codes*, Problemy Peredachi Informatsii, 18 (1982), pp. 7–13.

[4] ———, *A survey of superimposed code theory*, Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform., 12 (1983), pp. 229–242.

[5] K. Engel, *Sperner theory*, vol. 65 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1997.

[6] P. Erdős, P. Frankl, and Z. Füredi, *Families of finite sets in which no set is covered by the union of r others*, Israel J. Math., 51 (1985), pp. 79–89.

[7] Z. Füredi, *On r-cover-free families*, J. Combin. Theory Ser. A, 73 (1996), pp. 172–173.

[8] M. Ruszinkó, *On the upper bound of the size of the r-cover-free families*, J. Combin. Theory Ser. A, 66 (1994), pp. 302–310.

[9] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Math. Z., 27 (1928), pp. 544–548.