Lecturers: Hung Q. Ngo and Atri Rudra
Last update: October 11, 2011

Basic bounds for list disjunct and list separable matrices

1 Lower bounds

Recall that in the last lecture we have shown (in an exercise) that a (d, ℓ)-list-separable matrix is a $(d-1, \ell)$-list-disjunct matrix, and a (d, ℓ)-list-disjunct matrix is (d, ℓ)-list-separable. Hence, the optimal number of rows of a list-disjunct and a list-separable matrices are asymptotically the same. Thus, we shall only study the optimal number of rows of a list-disjunct matrices.

Let $t(d, \ell, N)$ denote the minimum number of rows of a (d, ℓ)-list-disjunct matrix with N columns. This lecture derives a couple of upper and lower bounds for this function.

Proposition 1.1 (Proposition 2 in [4])**.** . *Given positive integers $N \geq d + \ell$, we have*

$$t(d, \ell, N) \geq \log \binom{N}{d} - \log \binom{d + \ell - 1}{d}.$$

Exercise 1. Prove Proposition 1.1.

The following lower bound for (d, ℓ)-list-disjunct matrices is better than the similar bound proved in [3] in two ways: (1) the actual bounds are slightly better, and (2) the bound in [3] requires a precondition that $n > d^2/(4\ell)$ while ours does not. We make use of the argument from Erdős-Frankl-Füredi [5, 6], while [3] uses the argument from Ruszinkó [7] as presented in Alon-Asodi [1].

Lemma 1.2. *For any N, d, ℓ with $N \geq d + \ell$, we have*

$$t(d, \ell, N) > d \log \left(\frac{n}{d + \ell - 1} \right). \tag{1}$$

When $d \geq 2\ell$, the following bound holds

$$t(d, \ell, N) > \frac{\lfloor d/\ell \rfloor (d + 2 - \ell)}{2 \log \left(e \lfloor d/\ell \rfloor (d + 2 - \ell)/2 \right)} \log \left(\frac{N - d - 2\ell + 2}{\ell} \right). \tag{2}$$

Proof. Proposition 1.1 leads to (1) straightforwardly:

$$t(d, \ell, N) \geq \log \left(\frac{\binom{N}{d}}{\binom{d+\ell-1}{d}} \right) = \log \frac{N \cdots (N - d + 1)}{(d + \ell - 1) \cdots \ell} \geq \log \left(\frac{N}{d + \ell - 1} \right)^d = d \log \frac{N}{d + \ell - 1}.$$

Consider the case when $d \geq 2\ell$. Let $\mathbf{M}$ be a $t \times N$ (d, ℓ)-list-disjunct matrix. Fix a positive integer $w \leq t$ to be determined later. Let $\mathcal{C}$ denote the collection of all columns of $\mathbf{M}$, and think of $\mathcal{C}$ as a set family on $[t]$. Then, $\mathcal{C}$ satisfies the property that the union of any ℓ members of $\mathcal{C}$ is not covered by the union of any other

d members of $\mathcal{C}$. For any $C \in \mathcal{C}$, a subset $X \subseteq C$ is called a *private subset* of C if X is not a subset of any other C' in $\mathcal{C}$. Partition $\mathcal{C}$ into three sub-collections

$$\mathcal{C} = \mathcal{C}^{\text{p}}_{\geq w} \cup \mathcal{C}^{\text{np}}_{\geq w} \cup \mathcal{C}_{<w}$$

defined as follows.

$$\begin{aligned}
\mathcal{C}^{\text{p}}_{\geq w} &:= \{C \in \mathcal{C} \ : \ |C| \geq w \text{ and } C \text{ has a private } w\text{-subset}\} \\
\mathcal{C}^{\text{np}}_{\geq w} &:= \{C \in \mathcal{C} \ : \ |C| \geq w \text{ and } C \text{ has no private } w\text{-subset}\} \\
\mathcal{C}_{<w} &:= \{C \in \mathcal{C} \ : \ |C| < w\}.
\end{aligned}$$

We make three claims.

Claim 1. If $w \leq t/2$ then $|\mathcal{C}^{\text{p}}_{\geq w}| + \left\lfloor \frac{|\mathcal{C}_{<w}|}{\ell} \right\rfloor \leq \binom{t}{w}$.

Claim 2. Let $C_1, \cdots, C_\ell$ be any ℓ different members of $\mathcal{C}^{\text{np}}_{\geq w}$. For any integer $j \leq d/\ell - 1$ and any sub-collection $\mathcal{D} \subseteq \mathcal{C} \setminus \{C_1, \cdots, C_\ell\}$ such that $|\mathcal{D}| = j\ell$, we have

$$\left| \bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D \in \mathcal{D}} D \right| \geq (d - (j+1)\ell + 1)w + 1. \tag{3}$$

Claim 3. If $w \geq \frac{2(t - \lfloor d/\ell \rfloor)}{\lfloor d/\ell \rfloor (d+2-\ell)}$, then $|\mathcal{C}^{\text{np}}_{\geq w}| \leq d + \ell - 1$.

Let us complete the proof of the lemma before proving the claims. Set $w = \left\lceil \frac{2(t - \lfloor d/\ell \rfloor)}{\lfloor d/\ell \rfloor (d+2-\ell)} \right\rceil$. Then, $w \leq t/2$ when $d \geq 2\ell$. Note that $w < \bar{w} = \frac{2t}{\lfloor d/\ell \rfloor (d+2-\ell)}$ and the function $(te/w)^w$ is increasing in w when $w \in [0, t]$. From Claims 1 and 3,

$$\begin{aligned}
N = |\mathcal{C}| &= \left(|\mathcal{C}^{\text{p}}_{\geq w}| + |\mathcal{C}_{<w}| \right) + |\mathcal{C}^{\text{np}}_{\geq w}| \\
&\leq \left(\ell \left(|\mathcal{C}^{\text{p}}_{\geq w}| + \left\lfloor \frac{|\mathcal{C}_{<w}|}{\ell} \right\rfloor \right) + (\ell - 1) \right) + d + \ell - 1 \\
&\leq \ell \binom{t}{w} + d + 2\ell - 2 \\
&\leq \ell (te/w)^w + d + 2\ell - 2 \\
&\leq \ell (te/\bar{w})^{\bar{w}} + d + 2\ell - 2.
\end{aligned}$$

Inequality (2) follows.

We now prove Claim 1. Let $\mathcal{P}_1$ be a collection of private w-subsets of sets in $\mathcal{C}^{\text{p}}_{\geq w}$ such that $\mathcal{P}_1$ contains exactly one private w-subset per set in $\mathcal{C}^{\text{p}}_{\geq w}$. Let $\mathcal{L}$ be an arbitrary sub-collection of exactly ℓ different members of $\mathcal{C}_{<w}$, namely $\mathcal{L} \subseteq \mathcal{C}_{<w}$ and $|\mathcal{L}| = \ell$. Then, there must exist $C \in \mathcal{L}$ such that such that C is **not** a subset of any set in $\mathcal{P}_1 \cup \mathcal{C}_{<w} \setminus \mathcal{L}$. Otherwise, the union of sets in $\mathcal{L}$ will be covered by the union of at most $\ell \leq d$ sets in $\mathcal{C}$. We refer to such C as a *representative* of $\mathcal{L}$. For each $\mathcal{L}$, pick an arbitrary representative of $\mathcal{L}$ to be *the* representative of $\mathcal{L}$. Partition $\mathcal{C}_{<w}$ into $\left\lfloor \frac{|\mathcal{C}_{<w}|}{\ell} \right\rfloor$ sub-collections of cardinalities ℓ each, plus possibly one extra sub-collection whose size is less than ℓ. Let $\mathcal{P}_2$ be the set of the representatives of the first $\left\lfloor \frac{|\mathcal{C}_{<w}|}{\ell} \right\rfloor$ sub-collections. Then, $\mathcal{P}_1 \cup \mathcal{P}_2$ is a Sperner family, each of whose members is of cardinality at most w. For $w \leq t/2$, it is well-known (see, e.g., [2]) that $|\mathcal{P}_1 \cup \mathcal{P}_2| \leq \binom{t}{w}$. Noting that $|\mathcal{P}_2| = \left\lfloor \frac{|\mathcal{C}_{<w}|}{\ell} \right\rfloor$ and $|\mathcal{P}_1| = |\mathcal{C}^{\text{p}}_{\geq w}|$, Claim 1 follows.

Next, we prove Claim 2. Assume for the contrary that

$$\left|\bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D\in\mathcal{D}} D\right| \le (d-(j+1)\ell+1)w$$

for some $\mathcal{D}$ and j satisfying the conditions in the claim. For every $i \in [\ell]$, define

$$\begin{aligned} C'_i &:= C_i \setminus \bigcup_{D\in\mathcal{D}} D \cup C_1 \cdots \cup C_{i-1}. \\ x_i &:= \left\lfloor \frac{|C'_i|}{w} \right\rfloor \\ y_i &:= |C'_i| \bmod w. \end{aligned}$$

Then,

$$(d-(j+1)\ell+1)w \ge \left|\bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D\in\mathcal{D}} D\right| = \sum_{i=1}^{\ell} |C'_i| = \sum_{i=1}^{\ell}(x_i w + y_i) = w\left(\sum_{i=1}^{\ell} x_i\right) + \sum_{i=1}^{\ell} y_i.$$

Partition C'_i into x_i parts of size w each and one part of size $y_i \le w-1$. First, assume $\sum_{i=1}^{\ell} y_i > 0$, then $\sum_{i=1}^{\ell} x_i \le d-(j+1)\ell$. Because C_i has no private w-subset (and thus no private y_i-subset), the set C'_i can be covered by at most x_i+1 other sets in $\mathcal{C}$. The union $\bigcup_{i\in[\ell]} C'_i$ can be covered by at most $\sum_{i=1}^{\ell} x_i+\ell \le d-j\ell$ sets in $\mathcal{C}$. Those $d-j\ell$ sets covering the C'_i along with $j\ell$ sets in $\mathcal{D}$ cover the ℓ sets $C_i, i \in [\ell]$, which is a contradiction. Second, when $\sum_{i=1}^{\ell} y_i = 0$ we only need $\sum_{i=1}^{\ell} x_i \le d-(j+1)\ell+1 \le d-j\ell$ sets to cover the C'_i. The same contradiction is reached.

Finally we prove Claim 3. Suppose $|\mathcal{C}^{\text{np}}_{\ge w}| \ge d+\ell$. Consider $d+\ell$ sets $C_1,\dots,C_{d+\ell}$ in $\mathcal{C}^{\text{np}}_{\ge w}$. For $j = 0, 1, \cdots, \lfloor d/\ell \rfloor - 1$, define $\mathcal{D}_j = \{C_1, \cdots, C_{j\ell}\}$. ($\mathcal{D}_0 = \emptyset$.) Then, noting Claim 2, we have

$$\begin{aligned} t &\ge \bigcup_{i=1}^{d+\ell} C_i \\ &\ge \sum_{j=0}^{\lfloor d/\ell\rfloor-1} \left|\bigcup_{i=j\ell+1}^{(j+1)\ell} C_i \setminus \mathcal{D}_j\right| + \left|\bigcup_{i=d+1}^{d+\ell} C_i \setminus \bigcup_{i=1}^{d} C_i\right| \\ &\ge \sum_{j=0}^{\lfloor d/\ell\rfloor-1} \left[\big(d-(j+1)\ell+1\big)w+1\right]+1 \\ &= w\lfloor d/\ell\rfloor\left[d+1-\ell(\lfloor d/\ell\rfloor+1)/2\right] + \lfloor d/\ell\rfloor + 1 \\ &\ge \frac{1}{2} w \lfloor d/\ell\rfloor (d+2-\ell) + \lfloor d/\ell\rfloor + 1, \end{aligned}$$

which contradicts the assumption that $w \ge \frac{2(t-\lfloor d/\ell\rfloor)}{\lfloor d/\ell\rfloor(d+2-\ell)}$. □

2 Probabilistic upper bound and an application

Theorem 2.1. *Given positive integers $N \ge d+\ell$. Then,*

$$t(d,\ell,N) \le 2d\left(\frac{d}{\ell}+1\right)\left(\log\frac{N}{d+\ell}+1\right).$$

Proof. Fix positive integers n, q to be determined. Let $\mathbf{M}$ be the concatenation of the random code C_{out} and the identity code $C_{\text{in}} = \text{ID}_q$. The random code is of length n, each of whose codewords is chosen by setting each position to be a uniformly chosen symbol from an alphabet Σ of size q.

Suppose $\mathbf{M}$ is not (d,ℓ)-list-disjunct, then there exist two disjoint sets of columns S, T of $\mathbf{M}$ such that $S = \ell$ and $T = d$ such that the union of columns in S is contained in the union of columns in T. We call this pair (S,T) *bad for* $\mathbf{M}$. The columns in S and T correspond to two sets of codewords. Overloading notations, let S and T denote the two sets of codewords.

For each position $i \in [n]$, let T_i and S_i denote the set of symbols which the codewords in T and S have at that position, respectively. Then, the union of columns in S is contained in the union of columns in T if and only if for every position i we have $S_i \subseteq T_i$. For a fixed $i \in [n]$, the probability that $S_i \subseteq T_i$ is at most $(d/q)^\ell$. Hence, the probability that $S_i \subseteq T_i$ for all $i \in [n]$ is at most $(d/q)^{\ell n}$. Overall, the probability that a fixed pair (S,T) is bad for $\mathbf{M}$ ist at most $(d/q)^{\ell n}$.

Pick $n = 2\left(\frac{d}{\ell}+1\right)\left(\log\frac{N}{d+\ell}+1\right)$, $q = 3d \geq ed$, and taking the union bound over all choices of S and T, we obtain

$$\begin{aligned}
\text{Prob}[\mathbf{M} \text{ is not } (d,\ell)\text{-list-disjunct}] &= \text{Prob}[\text{some pair } (S,T) \text{ is bad for } \mathbf{M}] \\
&\leq \sum_{S,T} \text{Prob}[\text{the pair } (S,T) \text{ is bad for } \mathbf{M}] \\
&\leq \binom{N}{d+\ell}\binom{d+\ell}{\ell}(d/q)^{\ell n} \\
&\leq \exp\left((d+\ell)\ln\frac{Ne}{d+\ell} + \ell\ln\frac{(d+\ell)e}{\ell} - \ell n\right) \\
&< 1.
\end{aligned}$$

□

Corollary 2.2. *When $\ell = \Omega(d)$, we do have a nice reduction in the number of tests compared to the d-disjunct case:*

$$t(d, \Omega(d), N) = O(d\log(N/d)).$$

Corollary 2.3 (Optimal adaptive group testing)**.** *Consider the adaptive group testing problem where the tests are performed in stages: the next test can be designed after seeing the result of the previous test(s). Then, the optimal number of tests is $\Theta(d\log(N/d))$.*

Proof. For any adaptive group testing scheme, there are $\sum_{i=0}^{d}\binom{N}{i} = 2^{\Omega(d\log(N/d))}$ possible candidate sets of positives. With t tests we can only distinguish at most 2^t candidate positive sets. Hence, $t = \Omega(d\log(N/d))$. This type of argument is called the *information theoretic* reasoning.

A two stage group testing scheme with $O(d\log(N/d))$ tests can be designed as follows. We first use a (d,d)-list-disjunct matrix to identify a set of at most $2d-1$ items including all the positives. Then, an identity matrix of order $2d$ is used for the second stage to identify precisely the positives. □

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. De Bonis, L. Gąsieniec, and U. Vaccaro, *Optimal two-stage algorithms for group testing problems*, SIAM J. Comput., 34 (2005), pp. 1253–1270 (electronic).

[4] A. G. D′yachkov and V. V. Rykov, *A survey of superimposed code theory*, Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform., 12 (1983), pp. 229–242.

[5] 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.

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

[7] 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.