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 \((d, \ell)\)-list-separable matrix is a \((d-1, \ell)\)-list-disjunct matrix, and a \((d, \ell)\)-list-disjunct matrix is \((d, \ell)\)-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 \((d, \ell)\)-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*

\[
t(d, \ell, N) \geq \log \left( \frac{N}{d} \right) - \log \left( \frac{d + \ell - 1}{d} \right).
\]

**Exercise 1.** Prove Proposition 1.1.

The following lower bound for \((d, \ell)\)-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 \(N, d, \ell\) with \(N \geq d + \ell\), we have

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

(1)

When \(d \geq 2\ell\), the following bound holds

\[
t(d, \ell, N) > \frac{|d/\ell|(d + 2 - \ell)}{2 \log \left( e \left[ |d/\ell| (d + 2 - \ell)/2 \right] \right)} \log \left( \frac{N - d - 2\ell + 2}{\ell} \right).
\]

(2)

**Proof.** Proposition 1.1 leads to (1) straightforwardly:

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

Consider the case when \(d \geq 2\ell\). Let \(M\) be a \(t \times N\) \((d, \ell)\)-list-disjunct matrix. Fix a positive integer \(w \leq t\) to be determined later. Let \(\mathcal{C}\) denote the collection of all columns of \(M\), and think of \(\mathcal{C}\) as a set family on \([t]\). Then, \(\mathcal{C}\) satisfies the property that the union of any \(\ell\) members of \(\mathcal{C}\) is not covered by the union of any other
d members of \( C \). For any \( C \subseteq C \), a subset \( X \subseteq C \) is called a **private subset** of \( C \) if \( X \) is not a subset of any other \( C' \subseteq C \). Partition \( C \) into three sub-collections
\[
C = \mathcal{C}^p_{\geq w} \cup \mathcal{C}^{np}_{\geq w} \cup \mathcal{C}_{<w}
\]
defined as follows.
\[
\mathcal{C}^p_{\geq w} := \{ C \in C : |C| \geq w \text{ and } C \text{ has a private } w\text{-subset} \}
\]
\[
\mathcal{C}^{np}_{\geq w} := \{ C \in C : |C| \geq w \text{ and } C \text{ has no private } w\text{-subset} \}
\]
\[
\mathcal{C}_{<w} := \{ C \in C : |C| < w \}.
\]

We make three claims.

**Claim 1.** If \( w \leq t/2 \) then \( |\mathcal{C}^p_{\geq w}| + \left\lceil \frac{|\mathcal{C}_{<w}|}{t} \right\rceil \leq \binom{t}{w} \).

**Claim 2.** Let \( C_1, \ldots, C_\ell \) be any \( \ell \) different members of \( \mathcal{C}^{np}_{\geq w} \). For any integer \( j \leq d/\ell - 1 \) and any sub-collection \( D \subseteq C \setminus \{C_1, \ldots, C_\ell\} \) such that \( |D| = j\ell \), we have
\[
\left| \bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D \in D} D \right| \geq (d - (j + 1)\ell + 1)w + 1. \tag{3}
\]

**Claim 3.** If \( w \geq \frac{2(t-\lceil d/\ell \rceil)}{d/\ell + 2-\ell} \), then \( |\mathcal{C}^p_{\geq w}| \leq d + \ell - 1 \).

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

Inequality (2) follows.

We now prove Claim 1. Let \( \mathcal{P}_1 \) be a collection of private \( w \)-subsets of sets in \( \mathcal{C}^p_{\geq w} \) such that \( \mathcal{P}_1 \) contains exactly one private \( w \)-subset per set in \( \mathcal{C}^p_{\geq w} \). Let \( \mathcal{L} \) be an arbitrary sub-collection of exactly \( \ell \) different members of \( \mathcal{C}_{<w} \), namely \( \mathcal{L} \subseteq \mathcal{C}_{<w} \) and \( |\mathcal{L}| = \ell \). Then, there must exist \( C \subseteq \mathcal{C}_{<w} \) 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 \( \left\lceil |\mathcal{C}_{<w}|/\ell \right\rceil \) sub-collections of cardinalities \( \ell \) each, plus possibly one extra sub-collection whose size is less than \( \ell \). Let \( \mathcal{P}_2 \) be the set of the representatives of the first \( \left\lfloor |\mathcal{C}_{<w}|/\ell \right\rfloor \) 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 \( |\mathcal{P}_2| = \left\lfloor |\mathcal{C}_{<w}|/\ell \right\rfloor \) and \( |\mathcal{P}_1| = |\mathcal{C}^p_{\geq w}| \), Claim 1 follows.
Next, we prove Claim 2. Assume for the contrary that
\[ \left| \bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D \in \mathcal{D}} D \right| \leq (d - (j + 1)\ell + 1)w \]
for some \( \mathcal{D} \) and \( j \) satisfying the conditions in the claim. For every \( i \in [\ell] \), define
\[ C'_i := C_i \setminus \bigcup_{D \in \mathcal{D}} D \cup C_1 \cdots \cup C_{i-1}. \]
\[ x_i := \left\lfloor \frac{|C'_i|}{w} \right\rfloor \]
\[ y_i := |C'_i| \mod w. \]

Then,
\[ (d - (j + 1)\ell + 1)w \geq \left| \bigcup_{i=1}^{\ell} C_i \setminus \bigcup_{D \in \mathcal{D}} D \right| = \sum_{i=1}^{\ell} |C'_i| = \sum_{i=1}^{\ell} (x_iw + y_i) = w \left( \sum_{i=1}^{\ell} x_i \right) + \sum_{i=1}^{\ell} y_i. \]
Partition \( C'_i \) into \( x_i \) parts of size \( w \) each and one part of size \( y_i \leq w - 1 \). First, assume \( \sum_{i=1}^{\ell} y_i > 0 \), then \( \sum_{i=0}^{\ell} x_i \leq d - (j + 1)\ell \). 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 \( \sum_{i=1}^{\ell} x_i + \ell \leq d - j \ell \) sets in \( \mathcal{C} \). Those \( d - j \ell \) sets covering the \( C'_i \) along with \( j \ell \) sets in \( \mathcal{D} \) cover the \( \ell \) sets \( C_i, i \in [\ell], \) which is a contradiction. Second, when \( \sum_{i=1}^{\ell} y_i = 0 \) we only need \( \sum_{i=1}^{\ell} x_i \leq (d - (j + 1)\ell + 1) \leq d - j \ell \) sets to cover the \( C'_i \). The same contradiction is reached.

Finally we prove Claim 3. Suppose \( |\mathcal{C}_{[w]}| \geq d + \ell \). Consider \( d + \ell \) sets \( C_1, \ldots, C_{d+\ell} \) in \( \mathcal{C}_{[w]} \). For \( j = 0, 1, \ldots, \lfloor d/\ell \rfloor - 1 \), define \( \mathcal{D}_j = \{C_1, \ldots, C_{j\ell}\} \). (\( \mathcal{D}_0 = \emptyset \).) Then, noting Claim 2, we have
\[ t \geq \sum_{i=1}^{d+\ell} C_i \]
\[ \geq \sum_{j=0}^{\lfloor d/\ell \rfloor - 1} \left( \bigcup_{i=j\ell+1}^{(j+1)\ell} C_i \setminus \mathcal{D}_j \right) + \left| \bigcup_{i=d+1}^{d+\ell} C_i \setminus \bigcup_{i=1}^{d} C_i \right| \]
\[ \geq \sum_{j=0}^{\lfloor d/\ell \rfloor - 1} \left( (d - (j + 1)\ell + 1)w + 1 \right) + 1 \]
\[ = w[d/\ell] [d + 1 - \ell(\lfloor d/\ell \rfloor + 1)/2] + 1 \]
\[ \geq \frac{1}{2} w[d/\ell] (d + 2 - \ell) + [d/\ell] + 1, \]
which contradicts the assumption that \( w \geq \frac{2(t - \lfloor d/\ell \rfloor)}{[d/\ell](d + 2 - \ell)} \).

\[ \square \]

2 Probabilistic upper bound and an application

Theorem 2.1. Given positive integers \( N \geq d + \ell \). Then,
\[ t(d, \ell, N) \leq 2d \left( \frac{d}{\ell} + 1 \right) \left( \log \frac{N}{d + \ell} + 1 \right). \]
Proof. Fix positive integers $n, q$ to be determined. Let $M$ be the concatenation of the random code $C_{\text{out}}$ and the identity code $C_{\text{id}} = \text{ID}_q$. 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 $\Sigma$ of size $q$.

Suppose $M$ is not $(d, \ell)$-list-disjunct, then there exist two disjoint sets of columns $S, T$ of $M$ 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 $M$. 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 $M$ is 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

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

\[ \square \]

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. \[ \square \]

References


