GT Lecture 6

Review Macula's Theorem

$$\binom{n}{d}\left\{\overbrace{\begin{bmatrix} \\ M \text{ is } d\text{-disjunct} \\ \\ \end{bmatrix}}^{\binom{n}{k}}\right. \qquad k \geq d$$

Set $X = [n]$. Columns are subsets of X.

<u>Fact 1</u> If $d=1$, then M is 1-disjunct.

<u>Fact 2</u> Fix $d=1$ and $k>1$. How can we change the set of columns to increase the disjunctness of M. Clearly we need to chop off some columns to find M'. (<u>Combinatorial Designs</u>)

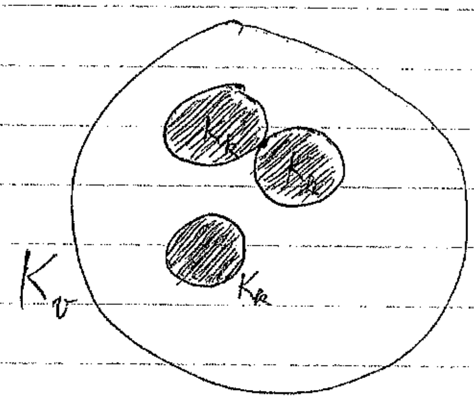
<u>Fact 3</u> If we let $x = \max\{|C_i \cap C_j| \mid C_i, C_j \text{ are two columns of } M\}$, and $|C_i| = k$ for each column C_i, then the disjunctness is (going to be) at least $\lceil \frac{k}{x} \rceil - 1$.

<u>Proof.</u> One column can cover the other column at most x elements. So, you need at least $\lceil \frac{k}{x} \rceil$ columns to cover another column. This implies that such a matrix M is $(\lceil \frac{k}{x} \rceil - 1)$-disjunct.

(*) If we can find a collection $\mathbb{B}$ of k-subsets in $[1, v]$ such that two subsets have at most one element in common, then we have an $M_{v \times |\mathbb{B}|}$ which is (k-1)-disjunct. (Use v in designs.)

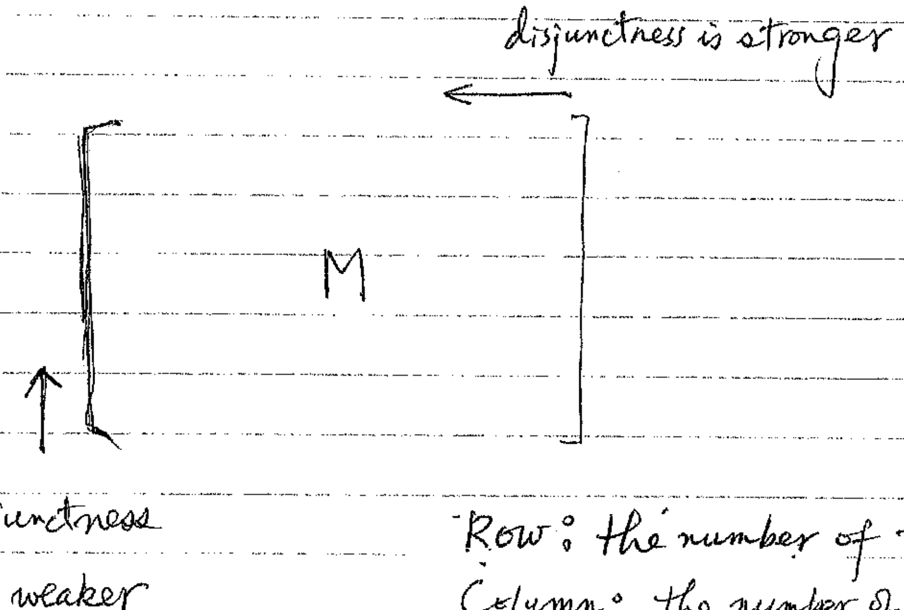
Fact 5 We can find at most $\left\lfloor \binom{v}{2} \Big/ \binom{k}{2} \right\rfloor$ such k-subsets such that any two of them contain at most one element in common.

Proof.



A k-subset of $[1, v]$ induces a k-clique and any two cliques contain at most one vertex in common.

So, we have the following observation.



Row: the number of tests
Column: the number of items

In a general setting, we can consider the following modification. Let $\underset{t\times n}{M}$ be a pooling design with fixed number of tests ~~t and n items to test.~~ Foreach column C_i, define $\varphi(C_i)$ be the number of minimum columns which can be used to cover C_i and $\varphi(C_i) = n$ if the union of $n-1$ other columns can not cover C_i. So, if M is d-disjunct, then $\varphi(C_i) \geq d+1$. Now, let

$$\varphi(\underset{t\times n}{M}) = \left(\sum_{i=1}^{n} \varphi(C_i)\right)\Big/ n .$$

Again, if M is d-disjunct, then $\varphi(M) \geq d+1$. But, the converse statement may not be true. Anyway, the following conclusion is indeed quite convicing: If $\varphi(M) \geq d+1$, then M is "almost very sure" d-disjunct.

(•) Question: If $\varphi(M) = 3.5$, then what can we say about the matrix M. Is M 2-disjunct? A quick conclusion is of course M is not 3-disjunct ($\varphi(M) \geq 4$ if it is 3-disjunct.)

(••) If we use such M ($\varphi(M) = 3.5$) to find three defective items, then what is the probability that the non-adaptive algorithm using M can successfully find them?

$$
M: \quad
\begin{array}{c}
\begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} \end{matrix} \\
\begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\end{array}
\begin{array}{l}
\\ \\ \\
(0) \to 1, 6, 9, 12 \\
(0) \to 2, 4, 9, 11 \\
\\
(0) \to 2, 6, 8, 10 \\
\\ \\ \\
\end{array}
$$

(•) Each column can be covered by three other columns but not two columns. $\sum_{i=1}^{12} d(C_i)/12 = 3$.

(•) M is 2-disjunct.

(•) Assume we have three defective items (3, 5, 7). Then, apply M, we obtain $\vec{y}^t = (1,1,0,0,1,0,1,1,1)$.

(*) By using the 0-pool tests, we conclude that $\{C_1, C_6, C_9, C_{12}\} \cup \{C_2, C_4, C_9, C_{11}\} \cup \{C_2, C_6, C_8, C_{10}\}$ is the set of negatives (good items) which is $\{C_1, C_2, C_4, C_6, C_8, C_9, C_{10}, C_{11}, C_{12}\}$. This is exactly the set of negatives and the others are positives.

(**) But, if the set of defective items is $\{1,2,3\}$, then the result will be different. (?) $\vec{y}^t = (1,1,1,1,1,1,1,1,1)$.

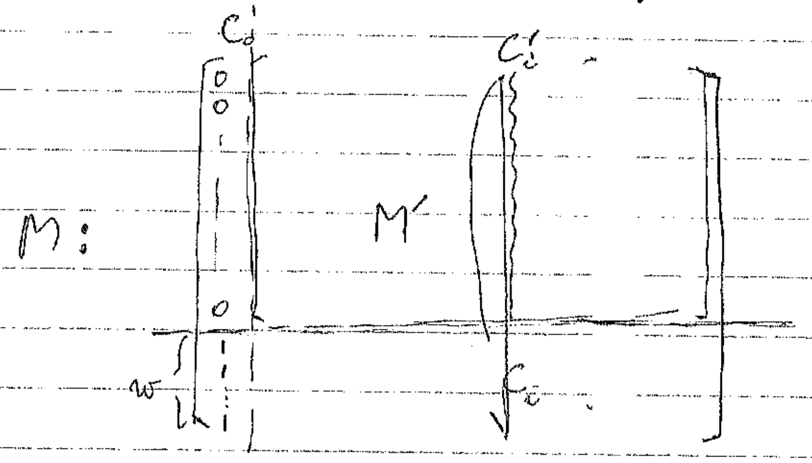
Problem What is the successful probability to find $d+1$ positive items from a set of n items by using a d-disjunct matrix? Here, we consider $\frac{n}{3} > d > 1$.

Definition

Let $t(d,n)$ denote the minimum number of rows (tests) for a d-disjunct matrix with n columns (items).

Fact 6. $t(d,n) \ge w + t(d-1,n-1)$ where $w = \max_{i=1}^{n} |C_i|$, C_i's are the columns.

Proof. Assume that the matrix with $t(d,n)$ rows is as follows (M):



Here, let $|C_0| = w$. Then M' is $(d-1)$-disjunct. Suppose not. In M', there exists one column $C_d' \subseteq C_1' \cup C_2' \cup \cdots \cup C_{d-1}'$. This implies that $C_d \subseteq \bigcup_{i=0}^{d-1} C_i$, a contradiction to the fact M is d-disjunct.

Hence, M' is a $(d-1)$-disjunct matrix with $n-1$ columns.

$t(d,n) - w \geq \underline{t(d-1, n-1)}$, since $t(d-1,n-1)$ is the minimum number of tests we need to find $d-1$ positives in $n-1$ items. ∎

<u>Remark</u> A good d-disjunct $t \times n$ matrix M does not contain a column which is too heavy (with larger weight) since it is easier to contain the union of the other columns.

<u>Problem</u> In a d-disjunct matrix $M_{t\times n}$, what is the distribution of column weights if $t \approx t(d,n)$?

<u>Fact 7.</u> $t(d,n) \geq \min\left\{\binom{d+2}{2}, n\right\}$.

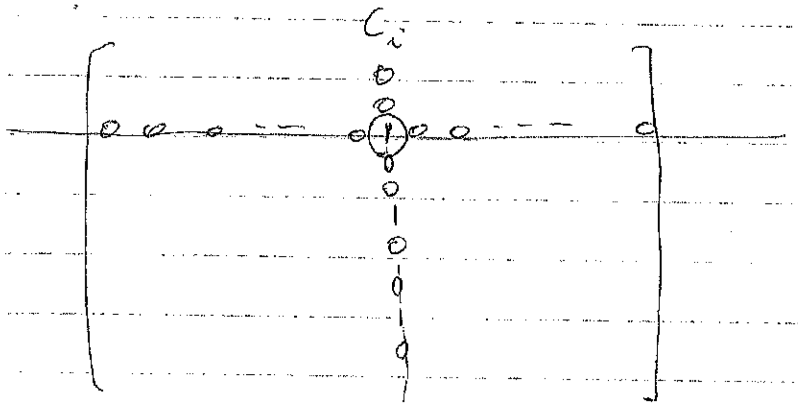
在 adaptive algorithm 最多问 $n-1$ 次，但是在 non adaptive 最多问 n 次。

<u>Proof.</u> By induction on n, and $n=1$ is true.

Let M be a d-disjunct $t(d,n) \times n$ matrix. First, we consider that M contains a column of weight $w \geq d+1$. By Fact 6,

$t(d,n) \geq d+1 + \underline{t(d-1, n-1)} \geq d+1 + \underbrace{\min\left\{\binom{d+1}{2}, n-1\right\}}_{\text{(by induction)}}$

$= \min\left\{d+1+\binom{d+1}{2}, d+1+n-1\right\} \geq \min\left\{\binom{d+2}{2}, n\right\}$. On the other hand, M contains all columns of weight at most d. Then, for each C_i (column) contains at least one element (isolated) which is not an element of all the other columns. For otherwise, C_i will be covered by d other columns, a contradiction.



Now, since every column has an element which does not occur in other columns, we need at least n tests to find the positives, i.e., $t(d,n) \geq n \geq \min\{\binom{d+2}{2}, n\}$. ∎

Theorem (Erdős, Frankl and Füredi, Israel J. Math., 1985, 79-89)

Let $c(w)$ be the number of columns with weight w. If $M_{t\times n}$ is a d-disjunct matrix, then $c(w) \leq \binom{t}{w} \Big/ \binom{w-1}{v-1}$ where $v = \lceil \frac{w}{d} \rceil$.

Proof. (Omitted.)

(*) Finding a suitable w is an important job in pooling e.g. $t=9$, $n=12$ and 2-disjunct : 9×12 matrix

$$c(3) = 12 \le \binom{9}{3} \Big/ \binom{2}{1} \quad \text{where} \quad v = \left\lceil \frac{3}{d} \right\rceil = 2, \; w = 3.$$

(•) This upper bound is in general "too big". Can you find a better one ?

(••) So far, the best : $t(d,n) \ge d(1+o(1))\log_2 n$. (?)

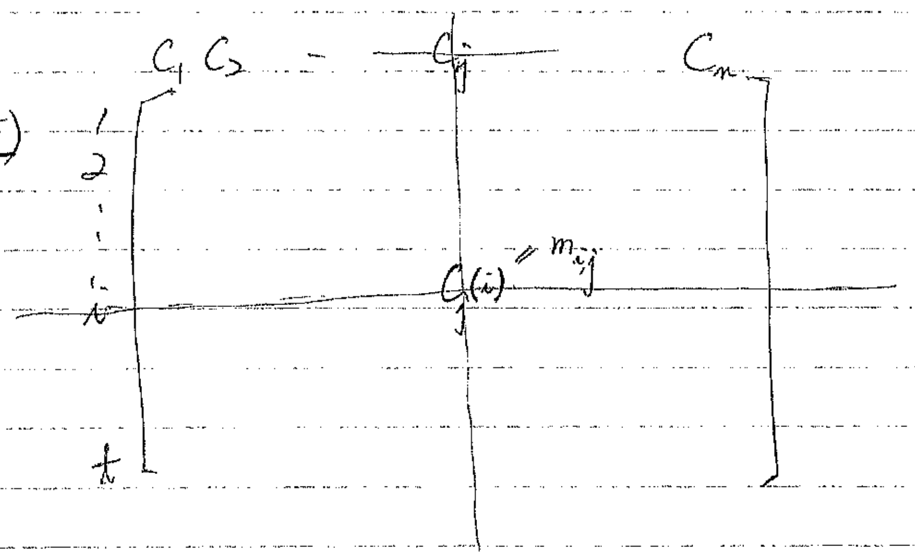
(*) $\bar{d}$-separable matrices do have "worse" decoding algorithm (in complexity), but comparing to d-disjunct matrices, we need less tests if the number of items is fixed to be n. That is, if $\bar{t}(d,n)$ denotes the number of tests in a $\bar{d}$-separable matrix with n items, then $\bar{t}(d,n) \le t(d,n)$.

(***) It is easier to extend a $(0,1)$-matrix M (pooling design) to a $[0,q-1]$-matrix M_q if separable matrices are concerned. Quite recently, disjunct matrices are also extended to q-case where $q \ge 3$.

We can extend the idea of d-separability (or $\bar{d}$-separability) to a $[0, q-1]$-matrix where $q \geq 2$. Regularly, q is taken as a prime power for the sake of computation. We start with an example.

Let M': $\begin{array}{c} \begin{matrix} c_1 & c_2 & c_3 & c_4 & c_5 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 2 \\ 1 & 2 & 3 & 1 & 3 \end{bmatrix} \end{array}$. Is M' 2-separable or $\bar{2}$-separable?

<u>Definition</u>
<u>(Descendant set)</u>



The matrix we consider is of the above form. We shall treat each column as a column vector (or a codeword) and thus $c_j(i)$ will be the ith coordinate of the vector c_j. Now, let $\mathbb{C}$ be a set of column vectors say $\{c_1, c_2, \cdots, c_d\}$. Then, we use $\mathbb{C}(j)$ to denote the set of jth coordinates of c_i's where $i = 1, 2, \cdots, d$, i.e., $\mathbb{C}(j) = \{c_1(j), c_2(j), \cdots, c_d(j)\}$. The descendant set (code) of $\mathbb{C}$ is $\mathbb{C}(1) \times \mathbb{C}(2) \times \cdots \times \mathbb{C}(t)$, denoted by $des(\mathbb{C})$.

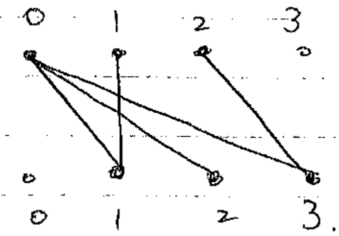
For example, in M' mentioned above, the descendant set of $\{c_1, c_2, c_4\} = \{0,1\} \times \{1,2\}$. (Note that c_1, c_2, c_4 are possible vectors of the form $(0,1)^t, (0,2)^t, (1,1)^t, (1,2)^t$ then.)

<u>Definition</u> M is d-separable if <u>for any two distinct d-subsets</u> (at most) for $\bar{d}$ of $\{c_1, c_2, \cdots, c_n\}$, $\mathbb{C}_1, \mathbb{C}_2,$ their corresponding descendant sets are distinct. $des(\mathbb{C}_1) \neq des(\mathbb{C}_2)$

Again, in M', all descendant sets of two column vectors are $\{0\}\times\{1,2\}$, $\{0\}\times\{1,3\}$, $\{0,1\}\times\{1\}$, $\{0,2\}\times\{1,3\}$, $\{0\}\times\{2,3\}$, $\{0,1\}\times\{1,2\}$, $\{0,2\}\times\{2,3\}$, $\{0,1\}\times\{1,3\}$, $\{0,2\}\times\{3\}$, $\{1,2\}\times\{1,3\}$. Since they are distinct, M' is 2-separable. Furthermore, M' is $\bar{2}$-separable. But, let $\mathbb{C}_1 = \{c_1, c_4, c_5\}$ and $\mathbb{C}_2 = \{c_3, c_4, c_5\}$. Then, $des(\mathbb{C}_1) = \{0,1,2\}\times\{1,3\}$ and $des(\mathbb{C}_2) = \{0,1,2\}\times\{1,3\}$. This implies that M' is not 3-separable.

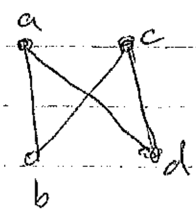
<u>Observation</u> Let M' be a $2\times n$ $[0, q-1]$-matrix. Then we can view each vector as an edge of a bipartite graph defined on (A, B) where $A = B = \mathbb{Z}_q$.

e.g. $M' \approx$



<u>Fact 1.</u> Let G be the graph induced by the columns of a $2 \times q$ $[0, q-1]$-matrix A. Then A is 2-separable if and only if G does not contain a 4-cycle.

<u>Proof.</u> ($\Rightarrow$) If G contains a 4-cycle (a, b, c, d), then $des(\{(a,b)^t, (c,d)^t\}) = des(\{(a,d)^t, (c,b)^t\})$. Hence A is not 2-separable. Since the reverse statement is also true, we have ($\Leftarrow$). ∎



<u>Fact 2.</u> Let C denote the set of column vectors. Then,
$\max |C| = ext.(n, n;\ C_4\text{-free})$.

<u>Definition</u> Let $z(m, n; h, k)$ denote the maximum number of edges in a bipartite graph defined on (A, B) where $|A| = m$ and $|B| = n$ which forbids $K_{h,k}$.

<u>Open problem</u> (Zarankiwicz's problem)

(∘) Most of n, $z(n,n;2,2)$ is unknown.

Now, we consider $\bar{3}$-separable case.

Observation If $\underset{2\times n}{A}$ is $\bar{3}$-separable, then its corresponding graph defined on (A,B) where $A=B=\mathbb{Z}_n$ can not contain a 4-cycle and a 6-cycle. That is the girth of G is at least 8.

Proof. If G contains a 6-cycle (a,b,c,d,e,f), then $des(\{a,b\},\{c,d\},\{e,f\}) = des(\{c,b\},\{e,d\},\{a,f\})$.

Fact 4. Let C be the set of column vectors of a $\bar{3}$-separable $2\times n$ matrix$\underset{A}{\wedge}$. Then $\max|C| = ext.(n,n; \underbrace{C_4, C_6\text{-free}}_{girth(G)\geq 8})$.

Open problem Let $\underset{(bipartite)}{G=(A,B)}$ such that $|A|=|B|=n$ and G is of girth 8. Find the maximum size of G.

(∘) Not much is known so far.

Open problem How about girth $2t$ for $t\geq 5$?

Open problem How to deal with $3\times n$ $[0,q-1]$-matrices?

Report topics

1. A study of fake coins problem
2. Determine $M(2,n)$.
3. Generalize "binary splitting algorithm"
4. A study of half-size induced subgraph of a bipartite graph problem
5. Use probabilistic method to construct disjunct matrices
6. Find the maximum size of a bipartite graph which is C_4-free
7. How to find a hidden subgraph of a complete graph.