GT Lecture 12

Date / / Dec. 28 – Jan. 4 No. 1

Direct Constructions of d-disjunct matrices

Review

Macula's Method

M: $\binom{n}{d} \times \binom{n}{k}$ matrix, indexed by $\binom{N}{d}$ and $\binom{N}{k}$ where $|N| = n$.

e.g. $N = 7$, $d = 2$, $k = 3$. $N = \mathbb{Z}_7$.

$$\begin{array}{c|ccccc} & 012 & 013 & 014 & 015 & \cdots \quad 567 \\ \hline 01 & 1 & 1 & & & \\ 02 & 1 & 0 & & & \\ 03 & 0 & 1 & & & \\ \vdots & \vdots & 0 & & & \\ & 0 & \vdots & & & \\ 12 & 1 & 0 & & & \\ 13 & 0 & 1 & & & \\ & & 0 & & & \\ & \vdots & \vdots & & & \\ 56 & 0 & \vdots & & & \end{array} \qquad \text{——— (1)}$$

$\binom{7}{2} \times \binom{7}{3}$
$= 21 \times 35$

$$\begin{cases} M(i,j) = 1 & \text{if the } i^{th}\ d\text{-subset of } N \text{ is contained in } j^{th}\ k\text{-subset of } N \\ M(i,j) = 0 & \text{otherwise} \end{cases}$$

$\Rightarrow$ M is d-disjunct. (So, the above pooling design is 2-disjunct)

Proof. Consider $d+1$ distinct columns, i.e., $d+1$ distinct k-subsets of N, $C_1, C_2, \cdots, C_{d+1}$. Since they are distinct

k-subsets of N, for each $j \in \{1, 2, \cdots, d\}$, $C_{d+1} \setminus C_j \neq \phi$, let $x_j \in C_{d+1} \setminus C_j$

Now, $D = \{x_1, x_2, \cdots, x_d\}$ is a multi-set, i.e., some element may occur more than once. Let D' be a subset (not a multiset) of D. Then, $D' \subseteq C_{d+1}$.

Moreover, $D' \subseteq \tilde{D}$ where $\tilde{D}$ is a d-subset of N occurred in the i^{th} row of M. Now, $\tilde{D} \nsubseteq C_j$ for each $j \in \{1, 2, \cdots, d\}$. Hence, $C_{d+1} \nsubseteq \bigcup_{j=1}^{d} C_j$. This concludes the proof. ∎

$$\begin{array}{cc} & \begin{array}{cccc|c} C_1 & C_2 & \cdots & C_d & C_{d+1} \; (\leq k\text{-subset}) \end{array} \\ \text{行名称为一个 } \tilde{D} & \left[\begin{array}{cccc|c} & & & & \\ 0 & 0 & \cdots & 0 & 1 \\ & & & & \end{array} \right] \end{array}$$

↑ There are $\binom{k}{d}$ 1's in this column!

Observation 1 Any two columns can have at most $\binom{k-1}{d}$ common 1's. (Any two columns are distinct!)

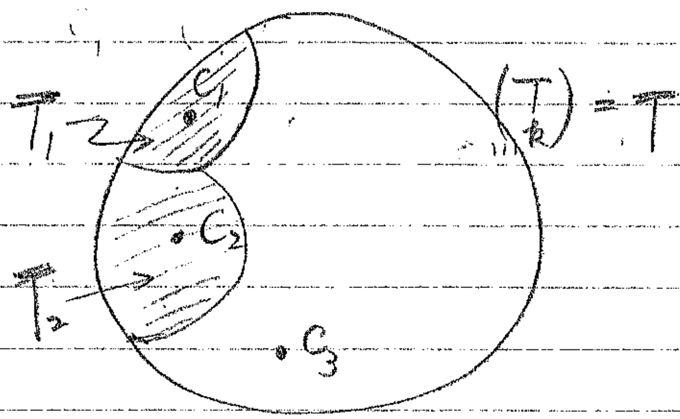
e.g. In (1), 015 and 014, do have a common one in row indexed by 01.

Observation 2 If we plan to cover a column by the other columns, then we need at least $\left\lceil \binom{k}{d} \Big/ \binom{k-1}{d} \right\rceil$ columns.

In general, this idea is called "Intersection Property" or column-intersection method.

<u>Algorithm</u>

Let $|T| = t$ and $\binom{T}{k} = \{k\text{-subsets of } T\}$. Define $r = \lceil t/(16d^2) \rceil$, $k = 4dr$ and $m = 4r$.



Choose C_1.

Choose $C_2 \in \mathcal{T} \setminus T_1$

$$T_1 = \{C \mid C \in \mathcal{T} \text{ and } |C \cap C_1| \ge m\}$$

$$T_2 = \{C \mid C \in \mathcal{T},\ C \notin T_1 \text{ and } |C \cap C_2| \ge m\}$$

$$\vdots$$

Choose $C_n \in \mathcal{T} \setminus \left(\bigcup_{i=1}^{n-1} T_i\right)$

$$T_n = \cdots$$

$$T_{n+1} = \emptyset$$

<u>Theorem</u> Let $C_1, C_2, \cdots, C_n$ be the columns and $1, 2, \cdots, t$ be the rows-indices. Then, we have a $t \times n$ d-disjunct matrix with $n \ge 2 \cdot 3^{\left(\frac{t}{16d^2} - 1\right)}$.

Proof. Since any two columns intersect in at most $m-1$ rows, (elements) to cover a column, it takes at least $\left\lceil \frac{k}{4r-1} \right\rceil$ columns which ($k=4dr$) is larger than d. Hence the pooling design is d-disjunct.

Now, we estimate n.

$$\text{For each } T_j,\ j=1,2,\cdots,n,\quad |T_j| \le \sum_{i=m}^{k} \boxed{\binom{k}{i}\cdot\binom{t-k}{k-i}} = \sum_{i=m}^{k} \boxed{b_i}$$

Not in common

common elements

This implies that $n \ge \dfrac{\binom{t}{k}}{\sum_{i=m}^{k} b_i}$.

Observe that, for $3r+1 \le i < k = 4dr$,

$$\frac{b_i}{b_{i-1}} = \binom{k}{i}\binom{t-k}{k-i} \Big/ \binom{k}{i-1}\binom{t-k}{k-i+1}$$

$$= \frac{\binom{k}{i}\cdot\binom{t-k}{k-i}}{\binom{k}{i-1}\cdot\binom{t-k}{k-i+1}}$$

$$= \frac{\dfrac{\cancel{k!}}{i!\,(k-i)!}}{\dfrac{\cancel{k!}}{(i-1)!\,(k-i+1)!}} \cdot \frac{\dfrac{\cancel{(t-k)!}}{(k-i)!\cdot(t-2k+i)!}}{\dfrac{\cancel{(t-k)!}}{(k-i+1)!\cdot(t-2k+i-1)!}}$$

$$= \frac{k-i+1}{i} \cdot \frac{(k-i+1)}{(t-2k+i)}$$

$$= \frac{(k-i+1)^2}{(i)(t-2k+i)} \le \frac{(4dr-3r)^2}{3r\cdot(16d^2r-8dr+3r)} \qquad (i=3r+1)$$

$$= \frac{(4d-3)^2}{3(16d^2-8d+3)} < \frac{1}{3}.$$

$$b_i < \left(\frac{1}{3}\right)^{i-3r} \cdot b_{3r}, \qquad \forall\ 3r+1 \le i < k = 4dr.$$

$$\sum_{i=m}^{k} b_i = \left(\frac{1}{3}\right)^{-3r} \cdot b_{3r} \cdot \sum_{i=m}^{k} \left(\frac{1}{3}\right)^{i}$$

$$= \left(\frac{1}{3}\right)^{-3r} \cdot b_{3r} \cdot \left(\frac{1}{3}\right)^{m} \cdot \left(1 + \frac{1}{3} + \cdots + \left(\frac{1}{3}\right)^{k-m-1}\right) \longrightarrow\ < \frac{1}{1-\frac{1}{3}}$$

$$< b_{3r} \cdot \left(\frac{1}{3}\right)^{m-3r-1} / 2 = 2^{-1} b_{3r} \cdot 3^{1-r}$$

Since $\binom{t}{k} = \sum_{i=0}^{k} b_i > b_{3r}$, $n \ge \left\lceil \binom{t}{k} \Big/ \sum_{i=m}^{k} b_i \right\rceil > 2 \cdot 3^{r-1} \ge 2 \cdot 3^{\left(\frac{t}{16d^2} - 1\right)}$.

Too small! (Not efficient!)

Can we choose a larger value? $b_{3r} = \binom{k}{i}\binom{t-k}{k-i}$ $(i = 3r)$

∘ Solve for t.

$$n \ge 2 \cdot 3^{\frac{t}{16d^2} - 1}$$

$$\log_3 n \ge \log_3 2 + \frac{t}{16d^2} - 1$$

$$t \le 16d^2 (\log_3 n - \log_3 2 + 1).$$

$$t(d,n) \le 16d^2 (\log_3 n - \log_3 2 + 1).$$

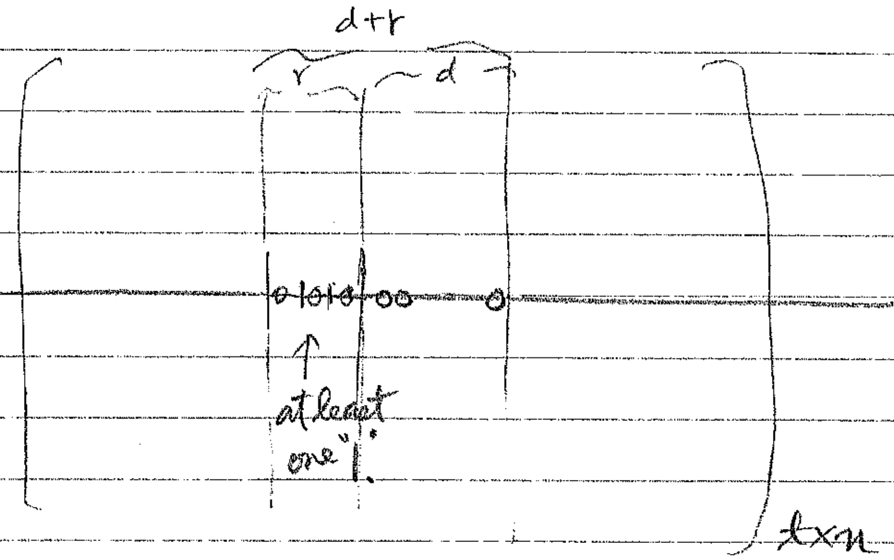
$$t(d,n) \le \underline{16d^2 (\log_3 2)} \underline{\log n (1 + o(1))}$$

(Non-constructive Method!)

With the help of designs, we can do better.

<u>Definition</u> A $t \times n$ binary matrix is (d,r)-disjunct if for $d+r$ arbitrary columns, the union of any r columns is not contained in the union of the other d columns.

There exists one such row!



D'yachkov and Rykov, A survey of superimposed code theory, Problems, Control Inform. Theory, 12 (1983), 1–13.

A $t \times n$ (d,r)-disjunct matrix is equivalent to the existence of a superimposed (d, n, r)-code of length t.

<u>Definition</u> (Superimposed code) t n d (positives)

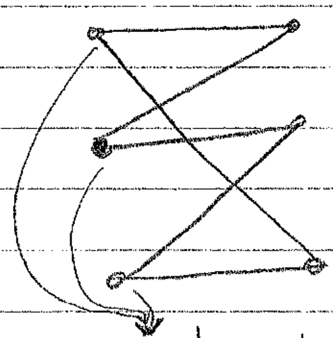
A superimposed code (binary) $SIC(n, N, m)$ consists of N binary codewords of length n, with the property that from the Boolean sum of any m-subset we are able uniquely determine the individual <u>codewords</u>

Construction of (d,r)-disjunct matrices

Definition An (ℓ, d, ε)-disperser is a bipartite graph $G = (A, B)$ satisfying the following conditions:

(1) $\forall\ S \subseteq A$ with $|S| \geq \ell$ is adjacent to at least $(1-\varepsilon)|B|$ vertices of B; and
$(|N(S)| \geq (1-\varepsilon)|B|)$

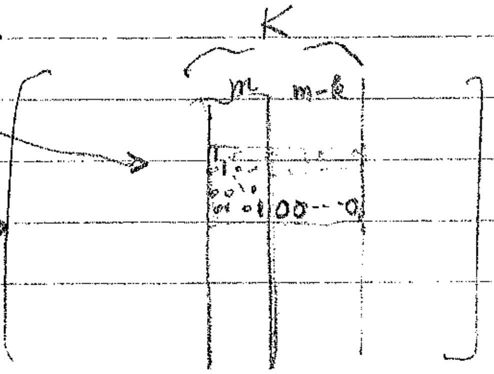
(2) $\forall\ v \in A$, $\deg_G(v) = d$.



$(2, 2, \frac{1}{10})$-disperser

Review

Definition A (k, m, n)-selector is a binary matrix with n columns such that for any k-set K of columns, at least m of them are isolated in K.

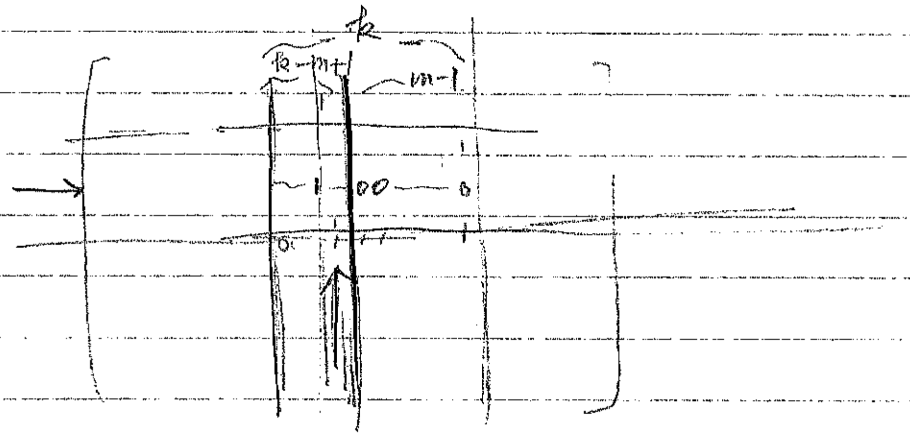


A column is called isolated if (in H) there exists a row incident to it but not to any other column.

Date / / No. 7

Fact A (k,m,n)-selector is $(m-1, k-m+1)$-disjunct.

Proof.



The union of any $k-m+1$ (K_1) columns is not contained in the union of $m-1$ (K_2) other columns since there exists a column (isolated) in K_1 which is not contained in the union of $m-1$ columns in K_2.
(In a (k,m,n)-selector, at least m out of k columns are isolated.)

(∘) Note, if $k-m+1=1$, i.e. $k=m$, then the matrix is in fact a $(k-1)$-disjunct matrix. Therefore, in any k columns, one of them is not contained in the union of the other $k-1$ columns.

So, consider $d=m-1$ and $r=k-m+1$. A (d,r)-disjunct matrix can be obtained by constructing a (k,m,n)-selector.

◎ B.S. Chlebus and D. R. Kowalski, Almost optimal explicit selector, LNCS 3623 (Springer-Verlag, 2005) 270-280.

(*) Let M be an $s \times n$ (h,h,n)-selector ($(h-1)$-disjunct)

and $H=(A,B)$ a (k,d,ε)-disperser.

$A = \mathbb{Z}_n$, $B = \mathbb{Z}_g' = \{0', 1', 2', \cdots, g-1'\}$ and rows of M are indexed

by $\mathbb{Z}_s$. Define an $sg \times n$ binary matrix M^* as follows:

$\forall$ $v \in A$ and $i = as+b$ where $a \geq 0$ and $0 \leq b < s$, $M^*(i,v) = 1$

if and only if v is neighbor of vertex a' in B and $M(b,v) = 1$.

Then, M^* is a (k,h,n)-selector provided $(1-\varepsilon)gh > kd$.
(See figure in next page!)

Proof. Suppose not. Then, there exists a k-set K of columns

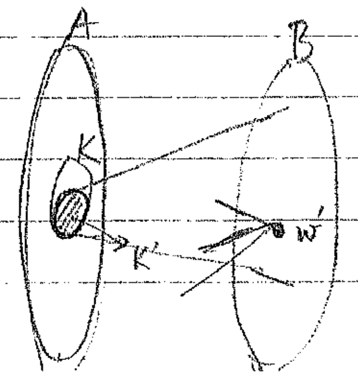
and a $(k-h+1)$-subset $K' \subseteq K$ such that every column in K' is

not isolated in K. (Since we need h isolated column, we have

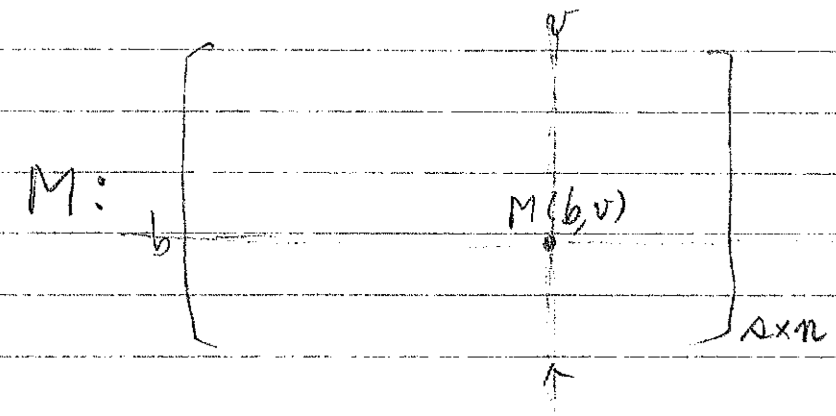
at most $k-h$ non-isolated columns in K.)

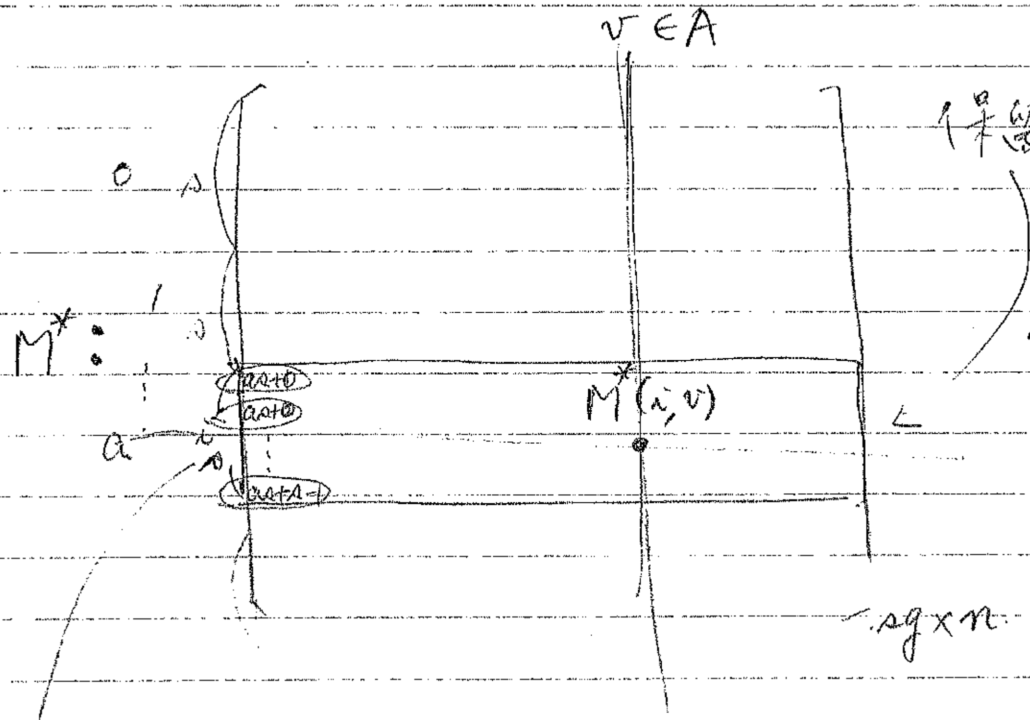
Now, consider the (k,d,ε)-disperser H. K' (and K) can be

"recognized" as a vertex subset of A. Let $w' \in B \cap N_H(K')$.



Claim:
~~w' has more than h~~
neighbors in K.
Suppose at most h.
Let $N_K(w') = \{v \in K, v \sim_H w'\}$.

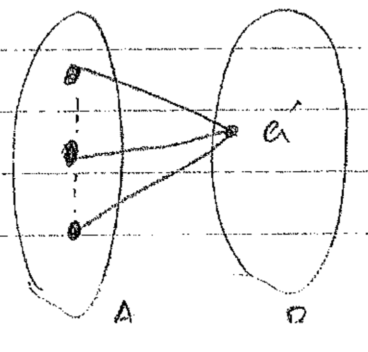




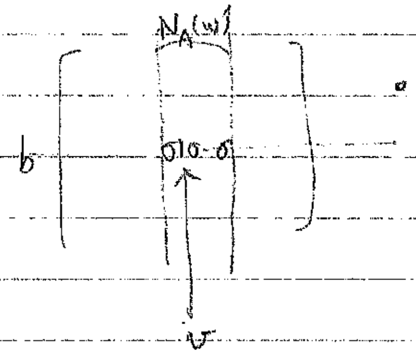
保留M中的某些行,那些行所對應的v要落在$N_H(a')$中。

$i = as + b$, $a \in \mathbb{Z}_g$ and $0 \le b < s$.

$M^*(i,v) = 1$ iff $M(b,v) = 1$ and $v \in N_H(a')$ where $a' \in B$ (corresponding to a).



Since M is an (h,h,n)-selector, there exists a row b in M such that



Now, the $(wA+b)$-th row of M^* intersects K with exactly one column (an isolated column), $\Rightarrow \in K'$.

Since H is a (k,d,ε)-disperser, $|N_H(K')| \geq (1-\varepsilon)|V_2| = (1-\varepsilon)g$.

Thus, $|\langle H, N_H(K')\rangle| \geq (1-\varepsilon)gh$. However, $\forall\ v \in A$, $\deg_H(v) = d$, we have at most dk edges which is less than $(1-\varepsilon)gh$ edges, a contradiction. ∎

More properties of a (d,r)-disjunct matrix.

① In a d-disjunct matrix, $t(d,1,n) \geq \log_2\binom{n}{d} - \log_2\binom{d+1}{d}$.

Proof.