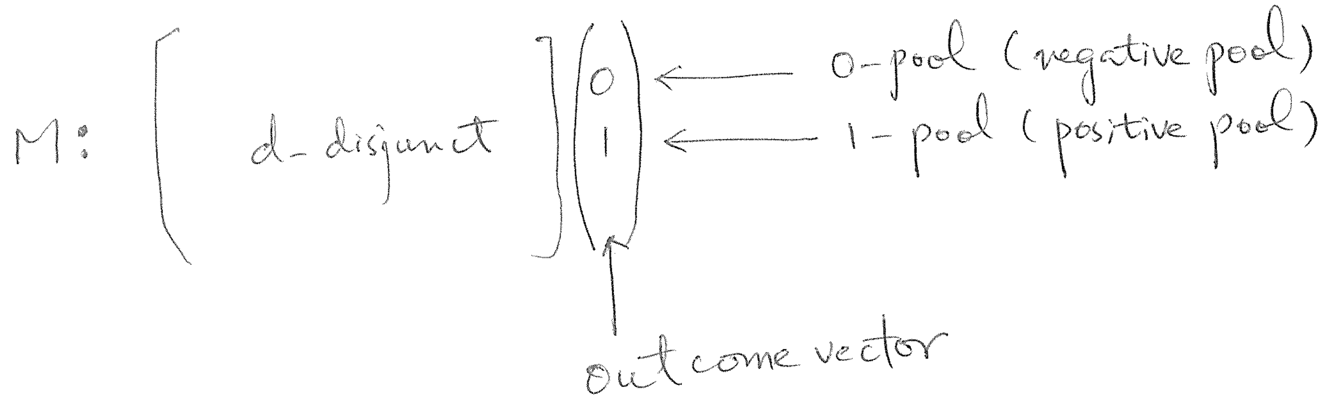
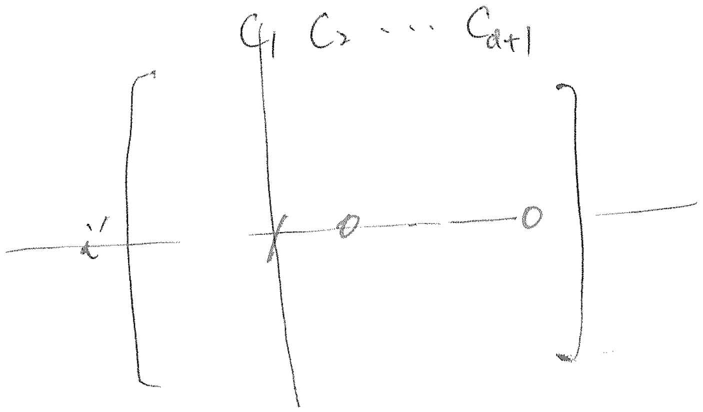


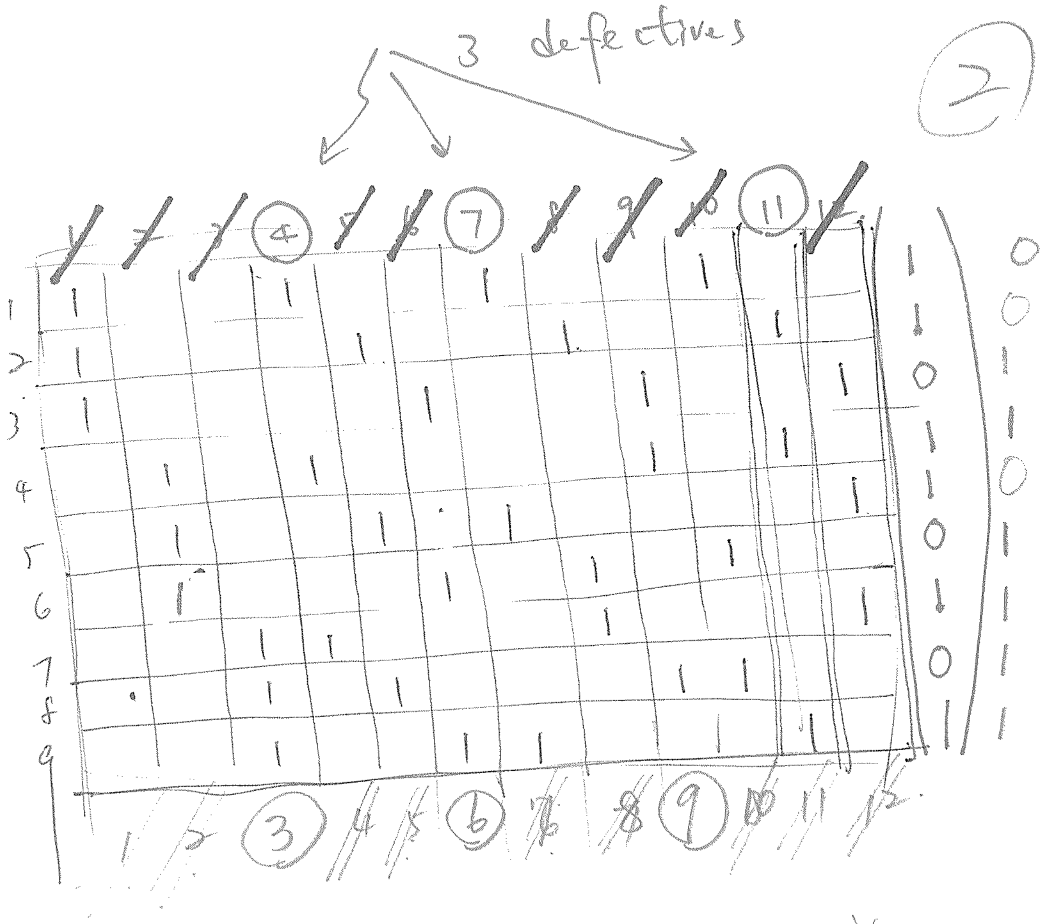
Observation

Let $c_1, c_2, \cdots, c_{d+1}$ be any $d+1$ columns of M.

Since $c_1 \nsubseteq \bigcup_{i=2}^{d+1} c_i$, in c_1, there exists an entry $m_{i',1}$ such that $m_{i',1}=1$ and $m_{i',i}=0$ for $i=2,\cdots,d+1$.



(*) 去掉 0-pools 所包含的 Items 之後，剩下的都是 defectives，而且不會超過 d 個；除非 defectives 的個數超過 d 個。



用一个 2-disjunct matrix 可以確定找出 3 個 defective items 嗎?

可以用於 2-stage Algorithms!

How to construct a d-disjunct matrix?

Theorem (A. J. Macula, DM 1996)

Let $M(d,k,n)$ be a (d,k)-containment matrix in $[1,n]$. Then $M(d,k,n)$ is a d-disjunct matrix with $\binom{n}{d}$ rows and $\binom{n}{k}$ columns.

Proof.

Let $C_{j_0}, C_{j_1}, \cdots, C_{j_d}$ be arbitraily selected distinct columns. Hence, $C_{j_0} \setminus C_{j_i} \neq \phi \ \forall\, i = 1,2,\cdots,d$. Let
(当成集合)
$x_i \in C_{j_0} \setminus C_{j_i}$, $i = 1,2,\cdots,d$ and $D_0 = \{ x_i \mid i = 1,2,\cdots,d \}$.

Clearly, $D_0 \subseteq C_{j_0}$. Let D $\subseteq C_{j_0}$ be a d-subset which contains D_0.

Thus, consider the t^{th} row corresponding to D, we have $m_{t,j_0} = 1$, but $m_{t,j_i} = 0 \ \forall\, i = 1,2,\cdots,d$. This implies that M is d-disjunct. ∎

$$M : \begin{array}{cc} & \text{k-subsets } C_j \\ \begin{array}{c}\text{d-subsets}\\ R_i\end{array} & \begin{bmatrix} & & \\ & m_{i,j} & \\ & & \end{bmatrix} \end{array}$$

$d < k$

$$m_{i,j} = \begin{cases} 1, & \text{if } R_i \subseteq C_j \\ 0, & \text{otherwise.} \end{cases}$$

1 – disjunct matrices

$$M: \quad \begin{array}{c|c} & C_1 \; C_2 \; \cdots \; C_n \\ \hline \begin{array}{c} 1 \\ 2 \\ \vdots \\ t \end{array} & \left(\begin{array}{c} \\ \\ \\ \\ \end{array} \right) \end{array}$$

We consider C_i is a subset of $[1, t]$ (instead of a column).

If M is 1-disjunct, then no C_i is contained in any other C_j, $j \neq i$. This implies that we are looking for a collection of subsets of $[1, t]$ such that for any two of them, A and B,

$A \not\subseteq B$ and $B \not\subseteq A$.

Now, let $T = [1, t]$. It is not difficult to see that $\binom{T}{\lfloor \frac{t}{2} \rfloor}$ is such a collection.

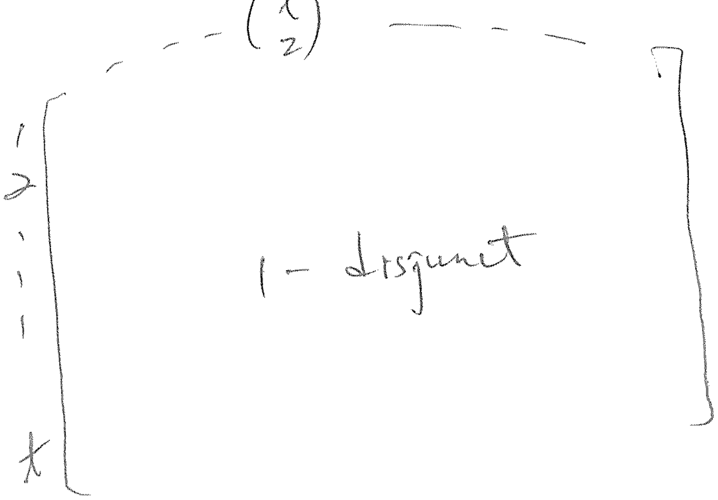
Question: Can we find more sets? That is can we find more than $\binom{t}{\lfloor \frac{t}{2} \rfloor}$ subsets of $[1, t]$ such that no two of them have the containment property?

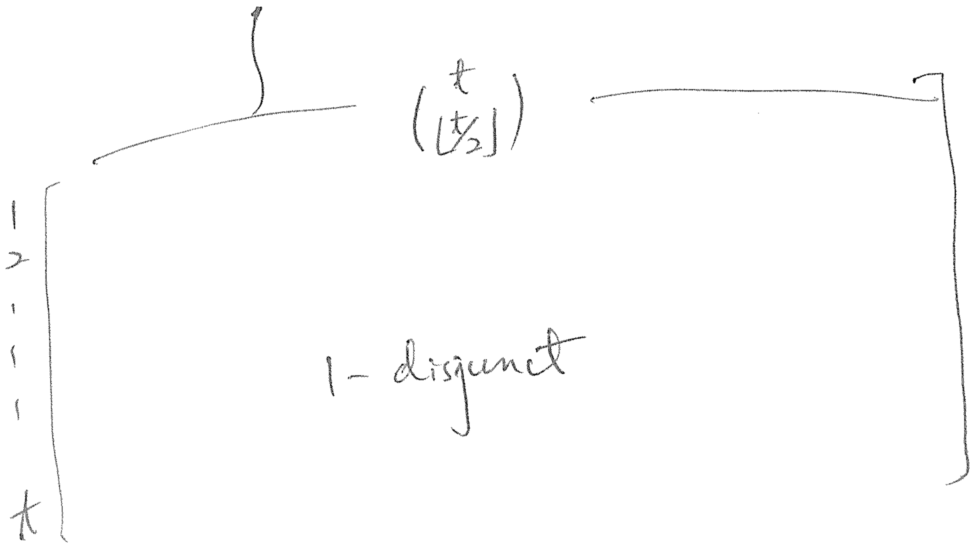
Observation

如果我们固定行的個數 n，设法加大 d，则列的個數自然会加大？

反过来，如果列的個數不增，则行数必定会变小。

從 Manla 的概念出發：





問題：

$$\begin{array}{cc} & \begin{matrix} C_1 & C_2 & \cdots & C_n \end{matrix} \\ & f(t) \\ \begin{matrix}1\\2\\\vdots\\t\end{matrix} & \begin{bmatrix} \quad & \quad & \quad & \quad \\ & & & \\ & & & \\ & & & \end{bmatrix}\end{array}$$

$f(t)$ 可能大於 $\binom{t}{\lfloor \frac{t}{2} \rfloor}$ 嗎？

如果 $1 \sim t$ 不变，d-disjunct 矩陣，$d > 1$，如何建構？

(Fact) Let $x = \max\{ |C_i \cap C_j| \mid C_i, C_j \text{ are two columns} \}$.

Moreover, let $|C_i| = k \ \forall i = 1, 2, \cdots, n$.

$d \geq \lceil \frac{k}{x} \rceil - 1$.

(Fact) If $x = 1$, then $d \geq k - 1$.

(Fact) Let $(X, \mathcal{B})$ be a <u>2-$(v, k, 1)$ design</u>, then the incidence matrix gives a $(k-1)$-disjunct matrix.

(Fact) Designs can be replaced by packing.

Some properties about the rows (tests) of a disjunct matrix.

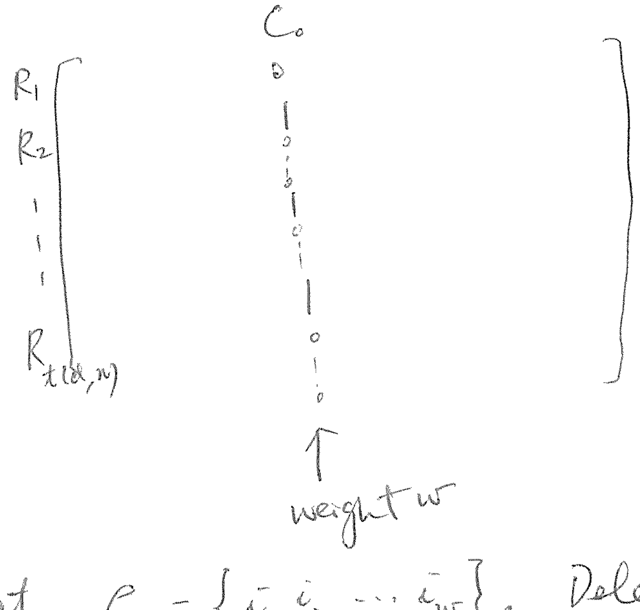
<u>Definition</u>

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

(*) Determining $t(d,n)$ is a very difficult problem for larger d and n.

<u>Lemma</u> $\quad t(d,n) \geq w + t(d-1, n-1)$ where $w = \max_{i=1}^{n} |C_i|$.

<u>Proof</u>: Let M be a matrix (d-disjunct) with n columns and $t(d,n)$ rows. Let C_0 be a column with weight w.



Assume that $C_0 = \{i_1, i_2, \cdots, i_w\}$. Deleting C_0 and R_{i_j}, $j = 1, 2, \cdots, w$, we have a matrix M' which has $t(d,n) - w$ rows and $n-1$ columns.

Claim M' is a $(d-1)$-disjunct matrix.

Suppose not. There exists a column C_1' in M' such that $C_1' \subseteq C_2' \cup C_3' \cup \cdots \cup C_d'$. This implies that

C_1 (with w rows back) $\subseteq C_0 \cup \underbrace{C_2 \cup \cdots \cup C_d}_{\text{with } w \text{ rows back}}$. M is not d-disjunct. (C_1 增加的在 C_0 中都有)。 $\rightarrow\leftarrow$ ▨

M' is $(d-1)$-disjunct. Hence

$t(d,n) - w \geq t(d-1, n-1)$ and we have the proof.
(minimum number of rows of a $(d-1)$-disjunct matrix)

Theorem $t(d,n) \geq \min \left\{ \binom{d+2}{2}, n \right\}$ for a d-disjunct matrix.

Proof. By induction on n. ($n=1$ is true).
Let M be a d-disjunct matrix with $t(d,n)$ rows and n columns.

Case 1 M has a column of weight $w \geq d+1$.

$$
\begin{aligned}
t(d,n) &\geq d+1+t(d-1,n-1) \\
&\geq d+1+\min\left\{ \binom{d+1}{2}, n-1 \right\} \\
&\geq \min\left\{ d+1+\binom{d+1}{2}, d+1+n-1 \right\} \\
&\geq \min\left\{ \binom{d+2}{2}, n \right\}.
\end{aligned}
$$

Case 2

M does not have a column of weight $\geq d+1$.

In this case, all columns are <u>isolated</u>. (If C_j is not isolated and $|C_j| \leq d$, then C_j is contained in the union of d other columns.) Since each column of weight $\leq d$ is isolated, $\sum_{w=1}^{d} r(w) \leq t$. This implies that

$$t(d,n) \geq \underbrace{\sum_{w=1}^{d} r(w) = n}_{\text{(all columns are isolated!)}} \geq \min\left\{\binom{d+2}{2}, n\right\}. \quad \blacksquare$$

Note

A subset of $\{1, 2, \cdots, t$ (stands for $t(d,n)$ here$)\}$ is called private if it is contained in one unique column. A column containing a private singleton subset is called isolated. We shall use $r(w)$ to denote the number of columns with weight w.

$$\begin{bmatrix} \\ \\ 0 \;—\; 0 \;\; 1 \;\; 0 \;—\; 0 \\ \\ \\ \end{bmatrix}$$

(**) If $\binom{d+2}{2} \geq n$, then $t(d,n) = n$.

(*) For each $d \leq n-1$ I_n is d-disjunct.

Theorem (Erdös, Frankl and Füredi, Israel J. Math. (1985) 79–89.)

Let M be a $t \times n$ d-disjunct matrix. Then

$$n(w) \leq \frac{\binom{t}{w}}{\binom{w-1}{v-1}} \quad \text{where } v = \left\lceil \frac{w}{d} \right\rceil .$$

---

e.g. $\left[\text{2-disjunct}\right]_{9\times 12}$. $\qquad n(3) = 12 \leq \dfrac{\binom{12}{3}}{\binom{2}{1}}$.

Theorem (Very good!)

$t(d,n) > d\,(1+o(1)) \log_2 n$.

Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Discrete Applied Mathematics 154 (2006) 1759–1762
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Note

A novel use of t-packings to construct d-disjunct matrices

Hung-Lin Fu, F.K. Hwang

Department of Applied Mathematics, National Chiao Tung University, Hsin Chu 30050, Taiwan, ROC

Received 16 March 2004; received in revised form 6 December 2005; accepted 14 March 2006
Available online 11 May 2006

Abstract

A t-packing is an ordered pair $(V, \mathbb{P})$ where V is a v-set and $\mathbb{P}$ is a collection of k-subsets (blocks) of V such that each t-subset of V occurs in at most one block of $\mathbb{P}$. If each t-subset of V occurs in exactly one block of $\mathbb{P}$, then $(V, \mathbb{P})$ is known as a Steiner (t, k, v)-design. In this paper, we explore a novel use of t-packings to construct d-disjunct matrices.

Keywords: t-packing; d-disjunct matrix

1. Introduction

In many DNA experiments, we want to know whether a clone (or a gene) contains a specific subsequence. If it does, we call the clone positive, otherwise, it is negative. Group testing is often used to identify the positive clones. A group test applies to an arbitrary subset of the clones and yields a positive outcome if and only if that subset contains a positive clone (the test does not reveal which or how many); otherwise, the outcome is negative. The goal is to use a minimum number of such tests to identify all positive clones. In biological applications, more important than the number of tests is the number of rounds in which these tests can be performed (all tests in the same round are performed parallelly). A 1-round test scheme is referred to as a "pooling design" in biological literature.

A major tool for the construction of pooling designs is the d-disjunct matrix. Let $M = (m_{ij})$ be a binary matrix. A column c_j of M can be viewed as a subset $\{i : m_{ij} = 1\}$ of the row index set $\{1, \dots, v\}$. M is d-disjunct if no column of M is contained in the union of any other d columns. Kautz and Singleton [4] proved that a d-disjunct matrix can identify all positive clones if their number does not exceed d. Further, there exists a simple decoding algorithm for positive clones.

A t-packing [1] is an ordered pair $(V, \mathbb{P})$ where V is a v-set and $\mathbb{P}$ is a collection of k-subsets (blocks) of V such that each t-subset of V occurs in at most one block of $\mathbb{P}$. If each t-subset of V occurs in exactly one block of $\mathbb{P}$, then $(V, \mathbb{P})$ is known as a Steiner (t, k, v)-design or a Steiner t-design denoted by $S(t, k, v)$.

t-packings (including Steiner t-designs) have been proposed [3,7] to construct d-disjunct matrices by using the blocks as columns and the elements as rows. For example, a block $\{x, y, z\}$ yields a column which has 1 in rows x, y, z and 0 in other rows. We say that two columns intersect at k rows if their inner product is k. By the definition of a t-packing, two columns intersect at at most $t-1$ rows. Hence it takes the union of at least $q = \lceil k/(t-1) \rceil$ columns, where k is the block size, to cover another column C. Thus the matrix is $(q-1)$-disjunct.

E-mail address: hlfu@math.nctu.edu.tw (H.-L. Fu).

0166-218X/$ - see front matter
doi:10.1016/j.dam.2006.03.009

In this paper, we propose a novel use of t-packings to construct d-disjunct matrices.

2. A new construction

Let $P(t,k,v)$ denote a t-packing with v elements and block size k. We show that we can use $P(t,k,v)$ to construct d-disjunct matrices for various d.

For each positive integer r, let M_r be a $\binom{v}{r} \times n$ binary matrix where the columns correspond to arbitrary n blocks from $P(t,k,v)$, the rows to all the $\binom{v}{r}$ r-subsets, and cell (i,j) is 1 if the r-subset corresponding to row i is a subset of the block corresponding to column j, otherwise cell (i,j) is 0.

Theorem 1. *M_r is a d-disjunct matrix with $d = \lceil \binom{k}{r} / \binom{t-1}{r} \rceil - 1$ for $1 \leqslant r \leqslant t-1$.*

Proof. Each column of M_r has $\binom{k}{r}$ r-subsets, i.e. $\binom{k}{r}$ rows with 1. Since any pair of blocks in $P(t,k,v)$ can take at most $t-1$ common elements, the corresponding pair of columns in M_r can intersect at at most $\binom{t-1}{r}$ rows. Thus in order to cover one column, at least $\lceil \binom{k}{r} / \binom{t-1}{r} \rceil$ other columns are needed. This concludes the proof. □

Note that the classical use of $P(t,k,v)$ corresponds to the special case $r=1$. Since $k \geqslant t$ in $P(t,k,v)$, it is easily verified:

Corollary 2. *d is maximized at $r = t-1$ where $d = \binom{k}{t-1} - 1$.*

It should be noted that larger r also implies more rows (tests). Macula [5] proposed the subset-containment construction of d-disjunct matrices, which is similar to our construction except that the columns of the matrix M are labelled by an arbitrary selection of n k-subsets of a set with v elements. Macula proved that M is r-disjunct. Our construction can be viewed as a modification of his construction when the column labels are judiciously selected, not arbitrarily. In our particular case, each column is labelled by a distinct block of $P(t,k,v)$, hence $n \leqslant \binom{v}{t}/\binom{k}{t}$, the number of blocks in $S(t,k,v)$. Note that our construction uses the same number $\binom{v}{r}$ of rows as Macula's, but increases d dramatically. The price paid is a smaller upper bound for n.

Note that a subset-containment design can be transformed to a classical pooling design where rows are indexed by numbers (or 1-subset) by simply matching the $\binom{v}{r}$ r-subsets to the set $\{1,\ldots,\binom{v}{r}\}$. Therefore the subset-containment construction (including ours) should simply be viewed as a new way to construct classical pooling designs.

3. An explicit construction

In this section, we shall use Steiner 3-designs to construct d-disjunct matrices $M_{m\times n}$ such that d is arbitrarily large and $m \leqslant n$. By Corollary 2, d is maximized at $r = t-1 = 2$, the number of rows $m = \binom{v}{2}$ and the number of columns $n = \binom{v}{3}/\binom{k}{3}$. Therefore, $n \geqslant m$ if and only if $2(v-2)/k(k-1)(k-2)$ is not less than 1.

The following result is essential to our explicit construction.

Theorem 3 (*Collbourn and Dinitz [2]*). *Let q be a prime power and $s \geqslant 2$. Then an $S(3, q+1, q^s+1)$ exists.*

The above Steiner designs are also known as spherical geometries.

By Theorem 1 and Corollary 2, we can use an $S(3,k,v)$ to construct a $(\binom{k}{2}-1)$-disjunct matrix $M_{m\times n}$, where $m = \binom{v}{2}$ and $n = \binom{v}{3}/\binom{k}{3}$. Then the following corollary can readily be stated from Theorem 3.

Corollary 4. *Let q be a prime power. For $d = \binom{q+1}{2} - 1$ and $s \geqslant 2$, there exists a d-disjunct matrix $M_{m\times n}$ where $n/m = 2(q^s-1)/(q+1)q(q-1)$.*

Remark. Clearly n/m is getting larger when s is increasing, in other words, $\lim_{s\to\infty} n/m = \infty$.

For $k \geqslant 4$, the existence and construction of $S(4, k, v)$ have been scattering in the literature. However, Rödl [6] proved that for v large, the number of columns in a t-packing can reach $(1 - o(1))\binom{v}{t}/\binom{k}{t}$. Thus even if a suitable $S(t, k, v)$ does not exist, for v large, using a t-packing can achieve almost as good a result.

4. Maximizing the number of columns

From Rödl's result, the number of blocks in a maximum t-packing is very close to $\binom{v}{t}/\binom{k}{t}$. In this section, our goal is to maximize the number of columns in a pooling design with fixed d and fixed number of rows. We will use the upper bound $\binom{v}{t}/\binom{k}{t}$ to represent the number of columns in a maximum t-packing and optimize over t. By our comments at the end of Section 2, we can transform our construction back to a classical (1-subset) design. Thus we have

Lemma 5. *The incidence matrix of a $P(t, k, v)$ such that $k = (t - 1)d + 1$ is a d-disjunct matrix.*

Proof. Since the rows of the matrix are indexed by V and any two blocks (columns) have at most $t - 1$ elements in common, the proof follows from Theorem 1. □

In the remaining part of this section, we consider $P(t, k, v)$ with $k = (t - 1)d + 1$ only.

Theorem 6. *For fixed v and d, the $P(t, k, v)$ which maximizes $f(t) = \binom{v}{t}/\binom{k}{t}$ occurs approximately when $t \approx (vc - 1)/(c + d)$ where $c = ((d - 1)/d)^{d-1}$.*

Proof. It is easily verified that $f(x)$ is concave on $[1, v]$. Thus $f(x)$ is maximized at $f'(x) = 0$, or approximately, $\binom{v}{x+1}/\binom{dx+1}{x+1} = \binom{v}{x}/\binom{dx-d+1}{x}$.

$$\frac{v!(dx-x)!}{(dx+1)!(v-x-1)!} = \frac{v!(dx-x-d+1)!}{(v-x)!(dx-d+1)!}$$

$$\frac{(v-x)(dx-d+1)!(dx-x)!}{(dx+1)!(dx-d-x+1)!} = 1$$

$$\frac{(v-x)(dx-x)(dx-x-1)\cdots(dx-d-x)}{(dx+1)(dx)\cdots(dx-d)} = 1$$

$$\frac{v-x}{dx+1}\cdot\frac{dx-x}{dx}\cdot\frac{((d-1)x-1)}{dx-1}\cdots\frac{(d-1)x-d}{dx-d} = 1$$

$$\frac{v-x}{dx+1}\cdot\frac{d-1}{d}\cdot\frac{d-1}{d}\cdots\frac{d-1}{d} > 1$$

$$\frac{v-x}{dx+1}\cdot\left(\frac{d-1}{d}\right)^{d-1} > 1$$

$$\frac{c(v-x)}{dx+1} > 1, \quad cv - cx > dx + 1$$

$$x < \frac{cv-1}{c+d}. \qquad \square$$

As an example, if $v = 100$ and $d = 5$, then $x < ((\frac{4}{5})^4 \cdot 100 - 1)/(\frac{4}{5})^4 + 5 = 39.96/5.4096 \approx 7.38$. Therefore, $x = 7$ may be the best choice for t to obtain as many columns as possible.

Not much is known about explicit t-packing when t is larger than 3. Hence, even if we know which is the best choice of t, construction of the corresponding t-packing is far from being settled. This observation also points out that the study of t-packing is an important topic in constructing d-disjunct matrices.

5. Conclusion

Using t-packings and using the containment method to construct d-disjunct matrices are both not new ideas. What we proposed is a hybridization of these two ideas to produce a dramatic increase of the value d. Using the same $S(v, k, t)$, the classical construction yields $d = \lceil k/(t-1) \rceil - 1$, while ours offers a spectrum of choices with $d = \lceil \binom{k}{r} / \binom{t-1}{r} \rceil - 1$, $1 \leqslant r \leqslant t-1$, and number of tests $= \binom{v}{r}$. Using the same number $\binom{v}{r}$ of tests, Macula's subset-containment construction yields $d = r$ while ours yields $d = \lceil \binom{k}{r} / \binom{t-1}{r} \rceil - 1$, paying a price of limiting the number of columns to $\binom{v}{t}/\binom{k}{t}$ instead of $\binom{v}{k}$. We also compare various choices of t to maximize the number of columns for fixed d and number of rows.

A referee suggested to us another type of comparison by fixing d and comparing the ratio of column number vs. row number. In particular, consider Macula's construction with $n = \binom{v}{t}$, $k = \binom{k}{t-1}$ and $r = \binom{k}{t-1} - 1$, which yields a d-disjunct matrix with $d = \binom{k}{t-1} - 1$ with the column/row ratio $=(n - k + 1)/k = (\binom{v}{t} - \binom{k}{t-1} + 1)/\binom{k}{t-1}$. On the other hand, by setting $r = t - 1$, then the construction in this paper also yields a d-disjunct matrix with the same $d = \binom{k}{t-1} - 1$ and the ratio $\binom{v}{t}/\binom{k}{t-1}\binom{v}{t-1}$, which is smaller than the first ratio in general.

Thus the comparison of the two constructions may depend on the particular need of each situation.

Acknowledgments

The authors wish to extend their gratitude to the referees for their helpful comments in revising this paper.

References

[1] A.E. Brouwer, Packing and covering of $\binom{k}{t}$-sets, Math. Center Tracts 106 (1979) 89–97.
[2] C.J. Collbourn, J.H. Dinitz (Eds.), The CRC Handbook of Combinatorial Designs, CRC Press, Boca Raton, FL, 1996, pp. 47–74.
[3] P. Erdös, P. Frankl, D. Füredi, Families of finite sets in which no set is covered by the union of r others, Israel J. Math. 51 (1985) 79–89.
[4] W.H. Kautz, R.R. Singleton, Nonrandom binary superimposed codes, IEEE Trans. Inform. Theory 10 (1964) 363–377.
[5] A.J. Macula, A simple construction of d-disjunct matrices with certain constant weights, Discrete Appl. Math. 80 (1996) 217–222.
[6] V. Rödl, On a packing and covering problem, European J. Combin. 5 (1984) 69–78.
[7] G.M. Sinha, H. Pesotan, B.L. Raktoc, Some results on t-complete designs, Ars Combin. 13 (1982) 195–201.