行政院國家科學委員會專題研究計畫 期中進度報告

關於組合設計理論及其相關應用的研究(1/3)

計畫類別：個別型計畫
計畫編號：NSC91-2115-M-009-012-
執行期間：91 年 08 月 01 日至 93 年 07 月 31 日
執行單位：國立交通大學應用數學系

計畫主持人：黃大原

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中 華 民 國 92 年 5 月 30 日

A Study of Dual Pairs of Set Systems

Tayuan Huang and Jiahsian Wang

Abstract: Each (0,1) matrix is associated with a dual pair of set systems. Some classical combinatorial structures including 2-designs, symmetric designs, quasi-symmetric designs are represented in this way. In terms of union or Boolean sums, some xondition were posed over a dual pair of set systems, so that it provides a way for pooling designs, i.e., non-adaptive group testing.

1. Preliminary

For a given (0,1)-matrix M of order $t \times n$, two set systems with ground sets $\{1,2,\ldots,n\}$ and $\{1,2,\ldots,t\}$ respectively, can be associated with rows as well as the columns of M naturally. Some combinatorial structures can be defined over these set systems by posing some conditions over those rows and columns of M. For example, the inner products among pairs of rows and pairs of columns were used in posing those conditions for 2-designs, symmetric designs and quasi-symmetric designs. On the other hand, Boolean sums among columns were used to define for disjunction, which lead to the notions of d-disjunct, d-separability and used for non-adaptive group-testing and pooling designs as well.

Let M be a $(0,1)$-matrix of order $t \times n$, let T_i be the subset of $\{1,2,\ldots,n\}$ with the i-th row of M as its characteristic vector, and C_j be the subset of $\{1,2,\ldots,n\}$ with the j-th row of M as its characteristic vector. We consider the following properties over the family $C = \{C_1, C_2, \ldots, C_n\} \subseteq 2^{[t]}$ and their dual families $T = \{T_1, T_2, \ldots, T_t\}$, where $C(D) = \bigcup_{i \in D} C_i$ for $D \in \binom{[n]}{d}$:

(i) $\left|C_i \cap C_j\right| < \frac{1}{d}\left|C_i\right| \;\; \forall i \neq j$,

(ii) $C \not\subset C(D)$ if $j \notin D, D \in \binom{[n]}{d}$,

(iii) $C(D) \neq C(D')$ wherever $D, D' \in \binom{[n]}{d}$ are distinct,

i.e., the d-term union are pairwise distinct.

Definition Let M be a $(0,1)$-matrix of order $t \times n$ and let C_j be the subset of $\{1,2,\ldots,n\}$ with the j-th row of M as its characteristic vector. Then the matrix M is called d-disjunct if $C = \{C_1,\ldots,C_n\} \subseteq 2^{[t]}$ satisfying the condition that

$C_j \not\subset \bigcup_{i\in D} C_i$.

2. d-disjunct

A d-disjunct matrix M of order $t \times n$, M provide a strategy for a non-adaptive group testing which can identify to d defects. Thus form the above discussion we have sufficient conditions for disjunctness of matrices : If

$\left|C_{i_0}\right| > \sum_{j=1}^{d} \left|C_{i_0} \cap C_{i_j}\right|$, since

$\left|C_{i_0} \cap \left(\bigcup_{j=1}^{d} C_{i_j}\right)\right| = \left|\bigcup_{j=1}^{d} C_{i_0} \cap C_{i_j}\right| \leq \sum_{j=1}^{d} \left|C_{i_0} \cap C_{i_j}\right| < \left|C_{i_0}\right|$, $C_{i_0} \not\subset \bigcup_{j=1}^{d} C_{i_j}$, then M is called d-disjunct.

In particular, let T_i be the test consisting of $\{j \mid M(i,j) = 1\}$, and

$\vec{r} = (x_1, x_2, \ldots, x_t)^t$ be the corresponding outcome vector. Suppose $D \subseteq [n]$ is the set of d defects, then $x_j = 1$ if and only of $j \in C(D)$. This group testing reports support$(\vec{r})$ as the set $C(D)$, consequent -ly, $j \in [n]$ is a defect if

$C_j \subseteq \text{support}(\vec{r}) = C(D)$.

Thus form the sufficient conditions for disjunctness of matrices, if M is a $(0,1)$- matrix of order $t \times n$ such that $JM = kJ$ and the off-diagonal entries of $M^t \cdot M$ are at most $\sim$. Let $d = \lceil k/\sim \rceil - 1$. For each $D = \{i_1, \ldots, i_d\} \subseteq [n]$, then

$\left|A_i \cap \left(\bigcup_{j=1}^{d} A_{i_j}\right)\right| \le \sum_{j=1}^{d} \left|A_i \cap A_{i_j}\right| \le d\sim \le k$. It follows that $A_i \not\subset \bigcup_{j=1}^{d} A_{i_j}$. This shows that M is a d-disjunct matrix of order $t \times n$. Therefore, an interesting problem in extremal set theory is the following: Find $F \subseteq \binom{[t]}{k} \subseteq 2^{[t]}$ with $|F|$ large as possible such that $|A \cap B| \le 1$ whenever $A, B \in F$ are different.

The following two theorems show that the relation between the family $C = \{C_1, C_2, \ \ldots, C_n\}$ and its dual family when they satisfied the above properties.

Theorem If $C = \{C_1, C_2, \ldots, C_n\} \subseteq 2^{[t]}$ satisfying the condition that $C(D) \ne C(D')$ for distinct $D, D' \in \binom{[n]}{d}$, then the dual family $T = \{T_1, T_2, \ldots, T_t\} \subseteq 2^{[n]}$ satisfying the following condition that for each $x \in \{0,1\}^t$, there exist $D \in \binom{[n]}{d}$ such that $|D \cap T_i| = 0$ if and only if $x_i = 0$.

Theorem Let $D \in \binom{[n]}{d}$, if $C = \{C_1, C_2, \ldots, C_n\} \subseteq 2^{[t]}$ satisfying the condition that $C_j \not\subset \bigcup_{i \in D} C_i$, then the dual family $T = \{T_1, T_2, \ldots, T_t\} \subseteq 2^{[n]}$ satisfying the following condition that $\bigcup_{j \notin C(D)} T_i = [n] - D$, and vice verse.

3. Some connections with combinatorial designs

A family $\{T_1, T_2, \ldots, T_t\} \subseteq 2^{[n]}$ is called a d-complete design if $\bigcup_{i \notin C(D)} T_i = [n] - D$ where $D \in \binom{[n]}{d}$. Note that if $i \notin C(D)$ then $M(i, \sim) = 0$ for all $\sim \in D$, i.e. $\sim \notin T_i$ for all $\sim \in D$. Hence $T_i \cap D = \varnothing$, i.e. $T_i \subseteq [n] - D$. For example,

1. $\{C_1, C_2, \ldots, C_b\}$ is a 2-$(v, k, 1)$ design, then $M_{v \times b}$ is $(k-1)$-disjunct, and hence $\{T_1, T_2, \ldots, T_v\}$ is $(k-1)$-complete.
2. If $\{C_1, C_2, \ldots, C_b\}$ is an affine resolvable 2-(v, k, a) design, then $|C_i \cap C_j| \le \frac{k^2}{v}$, hence $M_{v \times b}$ is $\left(\frac{k}{v} - 1\right)$-disjunct, and hence

$\{T_1, T_2, \ldots, T_v\}$ is $\left(\frac{k}{v} - 1\right)$-complete.

3. If $\{C_1, C_2, \ldots, C_b\}$ is any affine resolvable incomplete block design with

$|C_i| = k$ and $|C_i \cap C_j| \le q_1$ or q_2 then M is t-disjunct,

$t = \left\lfloor \frac{k}{\max\{q1, q2\}} \right\rfloor$, and hence $\{T_1, T_2, \ldots, T_v\}$ is t-complete.

Theorem If M is the point-block incidence matrix of a 2-$(v, k, 1)$ design and $D \in \binom{[n]}{d}$, then $D = [n] - \bigcup\{T_i \mid \text{for each } i \text{ with } M(i, D) = 0\}$.

Example :

A (16, 4, 1)-BIBD is presented by

{1 2 3 4, 5 6 7 8, 9 10 11 12, 13 14 15 16, 1 5 9 13, 2 8 10 15, 3 6 11 16, 4 7 12 14, 1 6 10 14, 2 7 9 16, 3 5 12 15, 4 8 11 13, 1 7 11 15, 2 6 12 13, 3 8 9 14, 4 5 10 16, 1 8 12 16, 2 5 11 14, 3 7 10 13, 4 6 9 15}.

The blocks of the dual incidence structure are as follows:

A_1 = {1,5,9,13,17} A_2 = {1,6,10,14,18} A_3 = {1,7,11,15,19} A_4 = {1,8,12,16,20}
A_5 = {2,5,11,16,18} A_6 = {2,7,9,14,20} A_7 = {2,8,10,13,19} A_8 = {2,6,12,15,17}
A_9 = {3,5,10,15,20} A_{10} = {3,6,9,16,19} A_{11} = {3,7,12,13,18} A_{12} = {3,8,11,14,17}
A_{13} = {4,5,12,14,19} A_{14} = {4,8,9,15,18} A_{15} = {4,6,11,13,20}
A_{16} = {4,7,10,16,17}.

Suppose we obtain the following result vector:

R(U) = (0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1).

When we carry out the algorithm identify with input *R(U)*, we compute the following:

j	M																			
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	**0**	1	1	1	**0**	1	1	1	**0**	1	1	1	**0**	1	1	1	**0**	1	1	1
3	0	1	1	1	0	1	**0**	1	0	1	**0**	1	0	1	**0**	1	0	1	**0**	1
4	0	1	1	1	0	1	0	**0**	0	1	0	**0**	0	1	0	**0**	0	1	0	**0**
6	0	**0**	1	1	0	1	0	0	0	1	0	0	0	**0**	0	0	0	1	0	0
10	0	0	**0**	1	0	**0**	0	0	0	1	0	0	0	0	0	0	0	1	0	0
12	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
15	0	0	0	**0**	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0

(Note that boxed entries are used to indicate when a “1” is changed to a “0”.) The positive set U is thus U = {10, 18}.

4 Some Classes Based on Ranked Posets

A few classes of non-adaptive pooling designs were proposed by Du and Ng, Macula and Weng and Huang . The first one is guaranteed to be *d*-error-detecting and thus $\lfloor d/2 \rfloor$-error-correcting, where *d*, a positive integer, is the maximum number of defectives (or positives). Hence, the number of errors which can be detected grows linearly with the number of positives. Also, this construction induces a construction of a binary code with maximum Hamming distance of at least $2d+2$. The second design is the *q*-analogue of a known construction on *d*-disjunct matrices.

For $1 \le d < k < n$, let $J(n,d,k)$ be the $(0,1)$-matrix of order $\binom{n}{d} \times \binom{n}{k}$ with rows and columns be indexed by respectively we define the matrix $J^{*}(n,d,k)$. Initially, we define the by letting its represented by the members of $\binom{[n]}{d}$ and $\binom{[n]}{k}$ respectively in the following way: for $D \in \binom{[n]}{d}$ and $K \in \binom{[n]}{k}$ the $(D,K)^{th}$ entry of the matrix $J(n,d,k)$ is 1 if and only if $D \subset K$. We then define $J^{*}(n,d,k)$ of order the $\left(\binom{n}{d}+n\right) \times \binom{n}{k}$ obtained by row augmenting the matrix $J(n,d,k)$ with $\not{c}^{\,c}(n,1,k)$.

For *q* being a prime power, let F_q denote the Galois field GF(*q*) of q elements. Let $\begin{bmatrix} F_q^m \\ l \end{bmatrix}$ denote the set of all *l*-dimensional subspaces (*l*-subspaces for short) of the *m*-dimensional vector space on F_q. For $m \ge k > d \ge 1$, Let $M_q(m,k,d)$ be the 0,1-matrix whose rows (resp. columns) are indexed by elements of $\begin{bmatrix} F_q^m \\ d \end{bmatrix}$ (resp. $\begin{bmatrix} F_q^m \\ k \end{bmatrix}$). We also order elements of these set lexicographically, i.e., the (i,j) entry of the matrix $M_q(m,k,d)$ is 1 if and only if the *i*-th *d*-subspace is a subspace of the *j*-th *k*-subspace of F_q^m.

Given integers $m \geq k > d \geq 1$. An *l-matching* is a matching of size *l* (i.e. it has *l* edges). Let *M*(*m*,*k*,*d*) be the (0,1)-matrix whose rows are indexed by the set of all *d*-matchings on K_{2m}, and whose columns are indexed by the set of all *k*-matchings on K_{2m}. All matchings are to be ordered lexicographically, i.e., *M*(*m*,*k*,*d*) is 1 if and only if the *i*-th *d*-matching is contained in the *j*-th *k*-matching.

Theorem For $m \geq k > d \geq 1$,

1. $M_q(m,k,d)$ is a *d*-disjunct matrix of order $v \times n$ with row weight $\begin{bmatrix} m-d \\ k-d \end{bmatrix}_q$ and column weight $\begin{bmatrix} k \\ d \end{bmatrix}_q$.
2. *M*(*m*,*k*,*d*) is a *d*-disjunct matrix of order $v \times n$ with row weight *g*(*m*-*d*,*k*-*d*) and column weight $\binom{k}{d}$.

Let *V* be an *n*-dimensional space over a finite field of order *q*. Let M be the incidence matrix with row entries indexed by the set of all 1-dimensional subspaces, columns entries indexed by the set of all *k*-dimensional subspaces of *V*, where $a_{i,j} = 1$ if and only if the *i*-th row is a 1-dimensional subspace contained in the *k*-dimensional subspace corresponding to the *j*-th column.

Theorem

1. Both $J(n,d,k)$ and $J^*(n,d,k)$ are d-disjunct; moreover

2. $d_H\left(B_d\left(J^*(n,d,k)\right)\right) \geq 4$ if $k-d \geq 3$.

Corollary Given integers $m > d \geq 1$, the following holds:
1. *M*(*m*, *k*, *d*) is *d*-error-detecting and $\lfloor d/2 \rfloor$-error-correcting；moreover
2. *M*(*m*, *k*, *d*) is (2*d*+1)-error-detecting and *d*- error-correcting if the number of positives

if known to be exactly d.

Theorem Let V be a vector space of dim n over a finite field $GF(q)$, and let M be a (0,1)-matrix row-indexed by $\begin{bmatrix} V \\ d \end{bmatrix}$, column-indexed by $\begin{bmatrix} V \\ k \end{bmatrix}$ such that $M(A,B)$=1 if $A \subseteq B$, and 0 otherwise. For a subset D of $\begin{bmatrix} V \\ k \end{bmatrix}$, let $L(D)$ be the Boolean sum of those columns corresponding to those k-subspaces in D. Then

$$e_m := \min_{D,D' \subseteq \begin{bmatrix} V \\ k \end{bmatrix}, |D|=|D'|=m} d_H(L(D), L(D')).$$

$$= \begin{cases} 2 \times \begin{bmatrix} k-m \\ d-m \end{bmatrix}_q, & \text{if } m \leq d, \\ 0, & \text{otherwise.} \end{cases}$$

5. References

[1] D.J. Balding, W.J. Bruno, E. Knill, D.C. Torney, A comparative survey of non-adaptive pooling designs, in: *Genetic Mapping and DNA Sequencing*, MN, 1994, Springer, New York, 1996, pp. 133-154.

[2] T. Beth, D. Jungnickel, H. Lenz, *Design Theory*, Cambridge University Press, Cambridge, 1986.

[3] T. Huang and K.S. Wang, Association Schemes and Group Testing with eror-tolerance, manuscript 2003.

[4] J.H. van Lint, R.M. Wilson, *A Course in Combinatorics*, Cambridge University Press, Cambridge, 1992.

[5] A.J. Macula, Error-correcting nonadaptive group testing with d^e-disjunct matrices, *Discrete Appl. Math.* 80 (2-3) (1997) 217-222.

[6] A. Macula, A simple construction of d-disjunct matrices with certain constant weights, to appear, *Discrete Math* (1994).

[7] H.Q. Ngo, D.-Z. Du, A survey on combinatorial group testing algorithms with applications to DNA , *DIMACS /DU2*, Amer. Math. Soc., Providence, RI, 2000.

[8] Weng, and Huang, Pooling Spaces and non-adaptive pooling designs, *Discrete Maths*, to appear.