J Comb Optim
DOI 10.1007/s10878-008-9179-4

A class of error-correcting pooling designs over complexes

Tayuan Huang · Kaishun Wang · Chih-Wen Weng

© Springer Science+Business Media, LLC 2008

Abstract As a generalization of d^e-disjunct matrices and $(w, r; d)$-cover-free-families, the notion of $(s, l)^e$-disjunct matrices is introduced for error-correcting pooling designs over complexes (or set pooling designs). We show that (w, r, d)-cover-free-families form a class of $(s, l)^e$-disjunct matrices. Moreover, a decoding algorithm for pooling designs based on $(s, l)^e$-disjunct matrices is considered.

Keywords Pooling design · Disjunct matrix · Decoding · Complex

1 Introduction

The notion of *superimposed code* was first introduced by Kautz and Singleton (1964) in the context of superimposed binary codes, and it was then generalized to d^e-*disjunct* matrices by D'yachkov et al. (1989) and by Macula (1997), to *superimposed* (s, l)-*code*, *superimposed* (s, l)-*design* by D'yachkov et al. (2002), and finally to $(w, r; d)$-*generalized-cover-free families* recently by Stinson and Wei (2004).

In the context of $(s, l; e)$-cover-free families, d-disjunct matrices with $(s, l; e) = (1, d; 1)$ have been generalized to d^e-disjunct matrices with $(s, l; e) = (1, d; e + 1)$ for error-correcting purpose (D'yachkov et al. 1989; Macula 1997); on the other hand, it has also been generalized to (s, l)-superimposed designs (D'yachkov et al. 2002) with $(s, l; e) = (s, l; 1)$ for the purpose of group testing over complexes. All these

This paper was presented in Algebraic Combinatorics-an international conference, held in Sendai, June 2006, in honor the 60th Birthday of Professor E. Bannai.

T. Huang · C.-W. Weng
Department of Applied Mathematics, National Chiao-Tung University, Hsinchu 30010, Taiwan

K. Wang (✉)
Sch. Math. Sci. & Lab. Math. Com. Sys., Beijing Normal University, Beijing 100875, China
e-mail: wangks@bnu.edu.cn

structures have found their applications in the designs of combinatorial group testing applicable to DNA library screening, and they are therefore called *pooling designs* with various additional properties.

More precisely, consider a set $[t] = \{1, 2, \ldots, t\}$ of molecules, the goal is to identify an unknown family $\wp = \{P_1, P_2, \ldots, P_k\}$ where the joint appearance of all molecules in each P_i causes a certain given biological phenomenon. An experiment, sometimes called a pool, can be applied to an arbitrary subset $S \subseteq [t]$ with two possible outcomes; a negative outcome implies S does not contain any $P_i \in \wp$, and a positive outcome implies otherwise. Members of $\wp$ are called positive complexes. Such a model is usually referred as the complex model. Of particular note is the basic assumption that members of $\wp$ are subject to non-inclusion. See (Du and Ngo 2000; Du and Hwang 2006) for more details and (Chen et al. 2008; Huang et al. 2007; Huang et al. 2008; Huang and Weng 2004) for related study.

In this paper, as a generalization of d^e-disjunct matrices and $(s, l; e)$-cover-free-families, the notion of $(s, l)^e$-*disjunct* matrices is introduced for error-correcting *pooling designs* (or called *set pooling designs, group testings over complexes*). We show that $(s, l; e)$-cover-free families form a class of $(s, l)^e$-disjunct matrices in Sect. 3; moreover, a decoding algorithm for error-correcting pooling designs based on $(s, l)^e$-disjunct matrices is given in Sect. 4.

2 Preliminary

For an $N \times t$ binary matrix M, let R_i and C_j denote the i-th row and j-th column of M, respectively. In this paper, we also let C_j denote the subset of $[N]$ consisting of all i with $M_{ij} = 1$. For positive integers s, l and t such that $s + l \le t$, let $\wp(s, l, t)$ be the family of all antichains $\wp = \{P_1, P_2, \ldots, P_k\}$ with $P_i \subseteq [t]$, $|P_i| \le l$, and $1 \le k \le s$. $\wp = \{P_1, P_2, \ldots, P_k\}$ is called an *antichain* if and only if P_i and P_j are not comparable whenever i and j are distinct.

The model of set pooling designs may be traced back to Torney (1999) and was carried out by D'yachkov et al. (2002).

Definition 2.1 (D'yachkov et al. 2002) A binary matrix M of order $N \times t$ is called

(1) a *superimposed* (s, l)-*code* if, for any two disjoint subsets S, L of $[t]$ with $|S| = s$ and $|L| = l$, there exists a row with entry 1 over L and 0 over S;
(2) a *superimposed* (s, l)-*design* if $\bigcup_{P_i \in \wp}(\bigcap_{j \in P_i} C_j) \ne \bigcup_{P'_i \in \wp'}(\bigcap_{j \in P'_i} C_j)$ for any two distinct $\wp = \{P_1, P_2, \ldots, P_k\}$, $\wp' = \{P'_1, P'_2, \ldots, P'_h\} \in \wp(s, l, t)$.

They showed that each (s, l)-superimposed code is an (s, l)-superimposed design, and each (s, l)-superimposed design is an $(s - 1, l)$-superimposed code and an $(s, l - 1)$-superimposed code as well. On the other hand, the following notion of $(w, r; d)$-cover-free families was introduced by Stinson and Wei (2004).

Definition 2.2 (Stinson and Wei 2004) Let w, r and d be positive integers. A set system $(X, \Im)$ is called a $(w, r; d)$-*cover-free-family* (or $(w, r; d) - CFF$) provided that, for any w blocks $B_1, \ldots, B_w \in \Im$ and any other r blocks $A_1, \ldots, A_r \in \Im$, we have that

$$\left| \bigcap_{1\le j\le w} B_i - \bigcup_{1\le j\le r} A_j \right| \ge d.$$

Note that the point-block incidence matrix of an $(l, s; 1)$-cover-free family is indeed a superimposed (s, l)-code. The notion of $(s, l)^e$-disjunct matrices is introduced as a common generalization of d^e-disjunct matrices and $(w, r; d)$-cover-free-family.

Definition 2.3 For positive integers s, l, t with $s + l \le t$, a binary matrix M of order $N \times t$ is called an $(s, l)^e$*-disjunct matrix* if

$$\left| \bigcap_{i\in A} C_i - \bigcup_{P_i\in\wp} \left(\bigcap_{j\in P_i} C_j \right) \right| \ge e$$

for any antichain $\wp = \{P_1, P_2, \ldots, P_k\} \in \wp(s, l, t)$, and for any $A \subseteq [t]$ with $|A| \le l$ and $A \notin \wp$.

An $(s, l)^e$-disjunct matrix M can be used for a pooling design in the following way: Let the columns of M be identified with the set of samples and its rows be identified with pools for testing such that $M(i, j) = 1$ if the j-th sample is included in the i-th pool. Suppose the set $[t] = \{1, 2, \ldots, t\}$ represents the set of samples with a (to be identified) positive family $\wp = \{P_1, P_2, \ldots, P_k\} \subseteq \wp([t])$, the power set of $[t]$, each test checks whether a pool contains at least one positive set $P_i \in \wp$ completely.

After the testing, the outcome vector

$$o(\wp) = o(\wp, M) = \text{the characteristic vector of the set } \bigcup_{P_i\in\wp} \left(\bigcap_{j\in P_i} C_j \right)$$

is reported for $\wp = \{P_1, P_2, \ldots, P_k\} \in \wp(s, l, t)$ if there is no error occurred during the processes, i.e., a test is reported *positive* only if it contains a certain positive subset P_i. Suppose instead that the report $o(\wp) + \epsilon$ with an error vector ϵ is received, Theorem 4.1 shows that the error occurring during the testing processing can be detected whenever the weight of ϵ is less than e, and the errors can be corrected whenever the weight of ϵ is no larger than $\lfloor \frac{e-1}{2} \rfloor$. In case $l = 1$, each $P_i \in \wp$ is reduced to a singleton, and it then reduces to k^e-disjunct matrices, whose decoding algorithm was discussed in (Huang and Weng 2003).

3 Some properties of $(s, l)^e$-disjunct matrices

Some good explicit constructions of generalized cover-free families, as well as nonconstructive existence results using the probabilistic method including the Lovasz Local Lemma can be found in (Stinson and Wei 2004), some bounds (i.e., necessary conditions) for generalized cover-free families were obtained through two different approaches. Theorem 3.1 shows that generalized cover free families provide a source of $(s,l)^e$-disjunct matrices. Some properties of $(s,l)^e$-disjunct matrices are given in Lemma 3.2 and Theorem 3.3 with the consideration of error tolerance over the pooling designs based on them.

Theorem 3.1 *The point-block incidence matrix M of an $(l,s;e)$-cover free family $\{C_1, C_2, \ldots, C_t\}$ is an $(s,l)^e$-disjunct matrix of order $N \times t$.*

Proof For any antichain $\wp = \{P_1, P_2, \ldots, P_k\} \in \wp(s,l,t)$, and for any $A \subseteq [t]$ with $|A| \le l$ and $A \notin \wp$, let $a_i \in P_i$ for $i \le k \le s$ and let $S \subseteq [t]$ be an s-subset containing $\{a_1, \ldots, a_k\}$. Then $\bigcup_{P_i \in \wp}(\bigcap_{j \in P_i} C_j) \subseteq \bigcup_{1 \le i \le k} C_{a_i} \subseteq \bigcup_{j \in S} C_j$, and hence

$$\left|\bigcap_{i \in A} C_i - \bigcup_{P_i \in \wp}\left(\bigcap_{j \in P_i} C_j\right)\right| \ge \left|\bigcap_{i \in A} C_i - \bigcup_{j \in S} C_j\right| \ge \left|\bigcap_{i \in A'} C_i - \bigcup_{j \in S} C_j\right| \ge e$$

where $A \subseteq A' \subseteq [t]$ with $|A'| = l$ because $\{C_1, C_2, \ldots, C_t\}$ is an $(l,s;e)$-generalized cover free family. □

Lemma 3.2 *Let M be an $(s,l)^e$-disjunct matrix, then $d_H(o(\wp), o(\wp')) \ge e$ whenever $\wp, \wp' \in \wp(s,l,t)$ are distinct.*

Proof Without loss of generality, we may assume that $\wp' - \wp$ is non-empty and $A \in \wp' - \wp$, we have

$$\left|\bigcap_{i \in A} C_i - \bigcup_{B \in \wp}\left(\bigcap_{j \in B} C_j\right)\right| \ge e$$

by definition, and therefore $d_H(o(\wp), o(\wp')) \ge e$. □

For an $(s,l)^e$-disjunct matrix M, we are interested to know the minimum distance, i.e., the minimum of the set $\{d_H(o(\wp), o(\wp')) | \wp, \wp' \in \wp(s,l,t)\}$.

Theorem 3.3 *Let M be an $(s,l)^e$-disjunct matrix. Given two distinct $\wp, \wp' \in \wp(s,l,t)$. Then the following hold*:

(1) *If $\wp \not\subset \wp'$ and $\wp' \not\subset \wp$, then $d_H(o(\wp), o(\wp')) \ge 2e$.*
(2) *If $\wp \subset \wp'$, then $d_H(o(\wp), o(\wp')) \ge e$.*

Proof (1) Let $\wp = \{P_1, P_2, \ldots, P_k\}$, $\wp' = \{P'_1, P'_2, \ldots, P'_h\} \in \wp(s,l,t)$. Then

$$\begin{aligned}
&d_H(o(\wp), o(\wp')) \\
&\quad = \left|\bigcup_{P_i \in \wp}\left(\bigcap_{j \in P_i} C_j\right) - \bigcup_{P'_i \in \wp'}\left(\bigcap_{j \in P_i} C_j\right)\right| + \left|\bigcup_{P'_i \in \wp'}\left(\bigcap_{j \in P'_i} C_j\right) - \bigcup_{P_i \in \wp}\left(\bigcap_{j \in P_i} C_j\right)\right| \\
&\quad \ge \left|\bigcap_{j \in P_i} C_j - \bigcup_{P'_i \in \wp'}\left(\bigcap_{j \in P_i} C_j\right)\right| + \left|\bigcap_{j \in P'_i} C_j - \bigcup_{P_i \in \wp}\left(\bigcap_{j \in P_i} C_j\right)\right| \\
&\quad \ge 2e.
\end{aligned}$$

The proof of (2) is similar to that of (1) and will be omitted. □

4 A decoding algorithm based on $(s, l)^e$-disjunct matrices

The methodology used by Kautz and Singleton (1964) has been generalized to a decoding method for pooling designs based on d^e-disjunct matrices (Huang and Weng 2003). In this section, we shall show that similar argument works well also for a decoding algorithm of pooling designs based on $(s, l)^e$-disjunct matrices.

Let χ_A with $A \subseteq [N]$ be the output vector for the group testing over the (to be identified) positive family $\wp = \{P_1, \ldots, P_k\}$. The following provides an decoding algorithm for the pooling design based on a $(s, l)^e$-disjunct matrix M.

Algorithm
Input: the output χ_A associated with $A \subseteq [N]$
Output: positive complexes $\wp$
$\wp_A := \emptyset$
While $Z \subseteq [N]$ and $|Z| \leq l$ do
If $|\bigcap_{j\in Z} C_j - \chi_A| \leq \lfloor \frac{e-1}{2} \rfloor$ then add Z into $\wp_A$
If $|\wp_A| > s$ or $\wp_A \neq \chi_A$ then output "there is an error"
else Output $\wp = \wp_A$

Theorem 4.1 *Let $A \subseteq [N]$, and let*

$$\wp_A = \left\{Z \mid |Z| \leq l \text{ and } |\textstyle\bigcap_{j\in Z} C_j - \chi_A| \leq \lfloor \frac{e-1}{2} \rfloor \right\}.$$

Then the following hold:

(1) *If $d_H(o(\wp), \chi_A) \leq \lfloor \frac{e-1}{2} \rfloor$, then $\wp = \wp_A$.*
(2) *Suppose $d_H(o(\wp), \chi_A) \leq e - 1$ and $|\wp_A| \leq s$. Then $o(\wp) = \chi_A$ if and only if $o(\wp_A) = \chi_A$.*

Proof (1) Since $\bigcap_{i\in Z} C_i \subseteq o(\wp)$ for any $Z \in \wp$, and then

$$\left|\bigcap_{j\in Z} C_j - \chi_A\right| \leq d_H(o(\wp), \chi_A) \leq \left\lfloor \frac{e-1}{2} \right\rfloor,$$

it follows that $Z \in \wp_A$. On the other hand, if $Z \in \wp_A$ but $Z \notin \wp$, then $|\bigcap_{j\in Z} C_j - o(\wp)| \geq e$ by definition. Since $d_H(o(\wp), \chi_A) \leq \lfloor \frac{e-1}{2} \rfloor$, we then have

$$\left|\bigcap_{j\in Z} C_j - \chi_A\right| \geq \left\lfloor \frac{e-1}{2} \right\rfloor + 1,$$

a contradiction.

(2) It is clear that if $\wp = \wp_A$. Now suppose that $\wp \neq \wp_A$. Then

$$d_H(o(\wp), \chi_A) > \left\lfloor \frac{e-1}{2} \right\rfloor$$

as just shown; in particular, $o(\wp) \neq \chi_A$. By Lemma 3.2,

$$d_H\big(o(\wp_A), \chi_A\big) \geq d_H\big(o(\wp), o(\wp_A)\big) - d_H\big(o(\wp), \chi_A\big) \geq e - (e-1) = 1,$$

and $o(\wp_A) \neq \chi_A$ as required. □

Acknowledgement The second author is supported by NSF of China.

References

Chen HB, Fu HL, Hwang FK (2008) An upper bound of the number of tests in pooling designs for the error-tolerant complex model. Opt Lett 2:425–431
D'yachkov AG, Rykov VV, Rashad AM (1989) Superimposed distance codes. Probl Control Inf Theory 18:237–250
D'yachkov AG, Vilenkin P, Macula AM, Torney D (2002) Families of finite sets in which no intersection of l sets is covered by the union of s others. J Comb Theory Ser A 99:195–218
Du D-Z, Hwang FK (2006) Pooling designs and nonadaptive group testing. World Scientific, Singapore
Du D-Z, Ngo HQ (2000) A Survey on Combinatorial Group Testing Algorithms with Applications to DNA Library Screening. DIMACS Ser Discrete Math Theor Comput Sci 55:171–182
Huang T, Weng C (2003) A note on decoding of superimposed codes. J Comb Optim 7:383–384
Huang T, Weng C (2004) Pooling spaces and non-adaptive pooling designs. Discrete Math 282:163–169
Huang H, Huang Y, Weng C (2007) More on pooling spaces. Discrete Math. doi:10.1016/j.disc.2007.11.073
Huang T, Wang K, Weng C (2008) Pooling spaces associated with finite geometry. Eur J Comb 29:1483–1491
Kautz W, Singleton R (1964) Nonrandom binary superimposed codes. IEEE Trans Inf Theory 10:363–377
Macula AJ (1997) Error-correcting nonadaptive group testing with d^e-disjunct matrices. Discrete Appl Math 80:217–222
Stinson DR, Wei R (2004) Generalized cover-free families. Discrete Math 279:463–477
Torney DC (1999) Sets pooling designs. Ann Comb 3:95–101