Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 155 (2007) 662–664
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Note

Exploring the missing link among d-separable, $\overline{d}$-separable and d-disjunct matrices

Hong-Bin Chen, Frank K. Hwang[1]

Department of Applied Mathematics, National Chiao Tung University, Hsinchu 300, Taiwan

Received 4 August 2004; received in revised form 16 May 2005; accepted 13 October 2006
Available online 21 December 2006

Abstract

d-Disjunct matrices, $\overline{d}$-separable matrices and d-separable matrices are well studied in various problems including group testing, coding, extremal set theory and, recently, DNA sequencing. The implications from the first two matrices to the last one are well documented. This paper gives an implication of the other direction for the first time.

Keywords: Pooling designs; Group testing; Separable; Disjunct

1. Introduction

Nonadaptive group testing, or *pooling designs* as in biology terminology, has been intensively studied recently due to its application to biological experiments (see [1,3] for general references). Three types of binary matrices have been found to be major tools in understanding and constructing pooling designs: we give their definitions. Consider a binary matrix M with strictly greater than d columns. Then M is:

d-separable if no two sets of d columns can have the same boolean sum,

$\overline{d}$*-separable* if no set of x columns can have the same boolean sum as another set of y columns if both x and y do not exceed d and

d-disjunct if no column $\leqslant$ the boolean sum of any other d columns.

Here $\leqslant$ is in the vector sense, i.e., for two vectors $V=(v_1, v_2, \ldots, v_t)$ and $V'=(v'_1, v'_2, \cdots, v'_t)$, $V \leqslant V'$ if and only if $v_i \leqslant v'_i$ for all $1 \leqslant i \leqslant t$.

These matrices have been studied elsewhere under other names. The d-separable matrix was first studied by Erdős and Moser [5] for $d=2$. Their problem is to determine the maximum number of columns a 2-separable matrix can have given t rows (the problem is still open). Frankl and Fűredi [6] called a d-separable matrix a *union-free hypergraph* by treating rows as vertices and columns as edges of a hypergraph (then the boolean sum of columns becomes the union of edges).

The $\overline{d}$-separable matrix was first studied by Kautz and Singleton [8] as a special kind of code named UD_d (uniquely decipherable code of order d). The d-disjunct matrix was also first studied by them under the name of ZFD_d

[1] This research was partially supported by the Republic of China, National Science Council Grant NSC 92-2115-M-009-014.
E-mail address: andan.am92g@nctu.edu.tw (H.-B. Chen).

0166-218X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2006.10.009

(zero-false-drop code of order d). The latter was also called a *d-cover-free family* by Erdős et al. [4], and called a *t-complete design* [2] if it is the incidence matrix of an incomplete block design.

Let n denote the number of columns in the given matrix. It is easily seen from definitions that d-separable ($\overline{d}$-separable, d-disjunct) implies k-separable ($\overline{k}$-separable, k-disjunct) for $1 \leqslant k < d < n$. The following relations are also well known [3]:

$$\overline{d+1}\text{-separable} \;\Rightarrow d\text{-disjunct} \;\Rightarrow \overline{d}\text{-separable} \;\Rightarrow d\text{-separable}.$$

In particular, d-disjunct $\Rightarrow \overline{d}$-separable with the option of dropping an arbitrary row.

Note that the relations between the three types of matrices miss a link from d-separable to k-disjunct or $\overline{k}$-separable for some $k < d$. In this paper we provide such a link, but not as strong as we like, i.e., k is not large enough. Therefore the value of our link is not in its practicality in constructing efficient k-disjunct or $\overline{k}$-separable matrices from known d-separable matrices, but rather in calling awareness to the existence of such a link, so that further research can improve on it.

2. The main results

Let $B(S)$ denote the boolean sum of a set S of columns.

Theorem 1. *Let M be a d-separable matrix. Then M is* $\overline{k+1}$*-separable*, $1 \leqslant k \leqslant d-1$, *if and only if M is k-disjunct.*

Proof. *Sufficiency*: Suppose to the contrary that there exist two distinct sets S and S' of columns in M, $|S| \leqslant k+1$, $|S'| \leqslant k+1$, such that $B(S) = B(S')$. By the d-separable property of M, we may assume $|S| < |S'| \leqslant k+1$. Then there exists a column $C \in S' \backslash S$. Since $C \leqslant B(S')$, we obtain $C \leqslant B(S)$, which violates the k-disjunct property of M.

Necessity:

Suppose M is not k-disjunct, i.e., there exist a column C and a set S of k other columns such that $C \leqslant B(S)$. Then $B(S) = B(S')$ where $S' = S \cup \{C\}$ and $|S|, |S'| \leqslant k+1$. Hence M is not $\overline{k+1}$-separable. □

We next give a construction showing how to convert a separable matrix to a disjunct matrix by adding tests and reducing d.

Theorem 2. *Let M be* $2d$*-separable. Then there exists a d-disjunct matrix by adding at most one row to M.*

Proof. If M is d-disjunct, then we are done. Suppose it is not. Then there exist a column C and a set S of d other columns such that $C \leqslant B(S)$. Add a row R which has a 1-entry at C and a 0-entry at each columns of S to break up $C \leqslant B(S)$ in M. Of course, there may exist C' and S', also with d columns, such that $C' \leqslant B(S')$ in M. Then we break it up by using R in the same fashion. However, what we need to show is that this procedure of setting the entries in R is not self-conflicting, i.e., there does not exist a column C such that $C \leqslant B(S)$, yet on the other hand $C \in S'$ while $B(S') \geqslant C' \neq C$ (since then C must have a 1-entry from $C \leqslant B(S)$, and a 0-entry from $B(S') \geqslant C'$).

Suppose to the contrary that there exist $C,\ C',\ S,\ S'$ as described above with $|S| \leqslant d$, $|S'| \leqslant d$ in M. Define

$$S_0 = \{C'\} \cup S \cup S',$$

$$S_1 = S_0 \backslash \{C\},$$

$$S_2 = S_0 \backslash \{C'\}.$$

Then

$$|S_0| \equiv s \leqslant 2d+1,$$

$$|S_1| = s-1 \leqslant 2d,$$

$$|S_2| = s-1 \leqslant 2d.$$

The fact $|S_1| = s - 1$ follows from $C \in S_0$, since $C \in S'$. Note that $S_1 \neq S_2$, but they have the same cardinality which is at most $2d$. We now show $B(S_1) = B(S_2)$, thus violating the assumption of $2d$-separability (which implies $(s-1)$-separability).

Since the only column in S_1 but not in S_2 is C', whose 1-entries are covered by S' which is in S_2, $B(S_1) \leqslant B(S_2)$. On the other hand, the only column in S_2 but not in S_1 is C, whose 1-entries are covered by S which is in S_1. Hence $B(S_2) \leqslant B(S_1)$. □

Theorem 3. *Let M be $2d$-separable. Then there exists a $\overline{d+1}$-separable matrix by adding at most one row to M.*

Proof. Theorem 3 follows from Theorems 2 and 1. □

3. Some impacts

A binary matrix M could be viewed as the incidence matrix of tests versus items, i.e., the rows are indexed by tests, the columns by items and the entry m_{ij} of cell (i, j) is 1 if test i contains item j, and 0 otherwise. Let $t(d, n)$ denote the minimum number of tests for a d-disjunct matrix with n items, and let $t_s(d, n)$ and $t_s(\overline{d}, n)$ denote the counterparts for d-separable and $\overline{d}$-separable matrix. It is well known [3] that

$$O(d^2 \log n/\log d) \leqslant t(d, n) \leqslant d^2 \log n.$$

Since d-disjunctness is stronger than d-separability or $\overline{d}$-separability, the upper bound of $t(d, n)$ remains an upper bound of $t_s(d, n)$ and $t_s(\overline{d}, n)$. However, the lower bound is not preserved. Currently, there is no good argument for lower bounds of $t_s(d, n)$ and $t_s(\overline{d}, n)$ except

$$t_s(d, n) \geqslant O(d \log n)$$

from the simple-minded argument that the number of distinct d-subsets, $\binom{n}{d}$, cannot exceed the number of distinct outcomes, 2^t. Theorem 3 shows that $t_s(d, n)$, $t_s(\overline{d}, n)$ and $t(d, n)$ have the same lower bound in the order of magnitude, i.e.:

Theorem 4. $t_s(\overline{d}, n) \geqslant t_s(d, n) \geqslant O(d^2 \log n/\log d)$.

In a sequential group-testing algorithm, the tests are done sequentially which means we can use outcomes from previous tests to determine what to test next. Let $t'(\overline{d}, n)$ or $t'(d, n)$ denote the minimum number of tests required to identify the $\overline{d}$ or d positive columns among n columns. Hwang et al. [7] proved:

Theorem 5. $t'(\overline{d}, n) - t'(d, n) \leqslant 1$.

It is easy to see that $t_s(\overline{1}, n) - t_s(1, n) \leqslant 1$, but otherwise, nothing is known about the difference. Setting $d = 1$ in Theorem 3, we obtain:

Corollary 6. $t_s(\overline{2}, n) - t_s(2, n) \leqslant 1$.

References

[1] D.J. Balding, W.J. Bruno, E. Knill, D.C. Torney, A comparative survey of nonadaptive pooling designs, in: Genetic Mapping and DNA Sequencing, IMA Volumes in Mathematics and Its Applications, Springer, Berlin, 1996, pp. 133–154.

[2] K.A. Bush, W.T. Federal, H. Pesotan, D. Raghavarao, New combinatorial designs and their applications to group testing, J. Statist. Planning Inference 10 (1984) 335–343.

[3] D.Z. Du, F.K. Hwang, Combinatorial Group Testing and Its Applications, second ed., World Scientific, Singapore, 2000.

[4] P. Erdős, P. Frankl, Z. Füredi, Family of finite sets in which no set is covered by the union of n others, Israel J. Math. 51 (1985) 79–89.

[5] P. Erdős, L. Moser, Problem 35, Proceedings on the Conference of Combinatorial Structures and their Applications, Calgary, 1969, Gordon and Breach, New York, 1970, p. 506.

[6] P. Frankl, Z. Füredi, Union-free hypergraphs and probability theory, European J. Combin. 5 (1984) 127–131.

[7] F.K. Hwang, T.T. Song, D.Z. Du, Hypergeometric and generalized hypergeometric group testing, SIAM J. Algebraic Discrete Methods 2 (1981) 426–428.

[8] W.H. Kautz, R.R. Singleton, Nonrandom binary superimposed codes, IEEE Trans. Inform. Theory 10 (1964) 363–377.