Journal of Combinatorial Theory, Series A 118 (2011) 2056–2058

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series A

www.elsevier.com/locate/jcta

A construction of pooling designs with surprisingly high degree of error correction

Jun Guo [a], Kaishun Wang [b]

[a] Math. and Inf. College, Langfang Teachers' College, Langfang 065000, China
[b] Sch. Math. Sci. & Lab. Math. Com. Sys., Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 3 November 2010
Available online 23 April 2011

Keywords:
Pooling design
Disjunct matrix
Error correction

ABSTRACT

It is well known that many famous pooling designs are constructed from mathematical structures by the "containment matrix" method. In this paper, we propose another method and obtain a family of pooling designs with surprisingly high degree of error correction based on a finite set. Given the numbers of items and pools, the error-tolerant property of our designs is much better than that of Macula's designs when the size of the set is large enough.

Pooling design is a mathematical tool to reduce the number of tests in DNA library screening [2–4]. A pooling design is usually represented by a binary matrix with columns indexed with items and rows indexed with pools. A cell (i, j) contains a 1-entry if and only if the ith pool contains the jth item. Biological experiments are notorious for producing erroneous outcomes. Therefore, it would be wise for pooling designs to allow some outcomes to be affected by errors. A binary matrix M is called *s^e-disjunct* if given any $s+1$ columns of M with one designated, there are $e+1$ rows with a 1 in the designated column and 0 in each of the other s columns. An s^0-disjunct matrix is also called *s-disjunct*. An s^e-disjunct matrix is called *fully s^e-disjunct* if it is not $s_1^{e_1}$-disjunct whenever $s_1 > s$ or $e_1 > e$. An s^e-disjunct matrix is $\lfloor e/2 \rfloor$-error-correcting (see [5]).

For positive integers $k \leqslant n$, let $[n] = \{1, 2, \ldots, n\}$ and $\binom{[n]}{k}$ be the set of all k-subsets of $[n]$.

Macula [10,11] proposed a novel way of constructing disjunct matrices by the containment relation of subsets in a finite set.

Definition 1. (See [10].) For positive integers $1 \leqslant d < k < n$, let $M(d, k, n)$ be the binary matrix with rows indexed with $\binom{[n]}{d}$ and columns indexed with $\binom{[n]}{k}$ such that $M(A, B) = 1$ if and only if $A \subseteq B$.

D'yachkov et al. [6] discussed the error-correcting property of $M(d, k, n)$.

E-mail addresses: guojun_lf@163.com (J. Guo), wangks@bnu.edu.cn (K. Wang).

0097-3165/$ – see front matter
doi:10.1016/j.jcta.2011.04.008

Theorem 1. *(See [6].) For positive integers* $1 \leqslant d < k < n$ *and* $s \leqslant d$*,* $M(d,k,n)$ *is fully* s^{e_1}*-disjunct, where* $e_1 = \binom{k-s}{d-s} - 1$.

Ngo and Du [13] constructed disjunct matrices by the containment relation of subspaces in a finite vector space. D'yachkov et al. [5] discussed the error-tolerant property of Ngo and Du's construction. Huang and Weng [9] introduced the comprehensive concept of pooling spaces, which is a significant addition to the general theory. Recently, many pooling designs have been constructed using the "containment matrix" method, see e.g. [1,7,8].

Next we shall introduce our construction.

Definition 2. Given integers $1 \leqslant d < k < n$ and $0 \leqslant i \leqslant d$. Let $M(i;d,k,n)$ be the binary matrix with rows indexed with $\binom{[n]}{d}$ and columns indexed with $\binom{[n]}{k}$ such that $M(A,B) = 1$ if and only if $|A \cap B| = i$.

Note that $M(i;d,k,n)$ and $M(d,k,n)$ have the same size, and $M(i;d,k,n)$ is an $\binom{n}{d} \times \binom{n}{k}$ matrix with row weight $\binom{d}{i}\binom{n-d}{k-i}$ and column weight $\binom{k}{i}\binom{n-k}{d-i}$. Since $M(d;d,k,n) = M(d,k,n)$, our construction is a generalization of Macula's matrix.

Let $B \in \binom{[n]}{k}$ and $C = [n] \setminus B$. Then, for any $D \in \binom{[n]}{d}$, $|D \cap B| = i$ if and only if $|D \cap C| = d - i$. Therefore, $M(i;d,k,n) = M(d-i;d,n-k,n)$ when $n > k + d - i$. Since $i \leqslant \lfloor d/2 \rfloor$ if and only if $d - i \geqslant \lfloor (d+1)/2 \rfloor$, we always assume that $i \geqslant \lfloor (d+1)/2 \rfloor$ in this case.

Theorem 2. *Let* $1 \leqslant s \leqslant i$*,* $\lfloor (d+1)/2 \rfloor \leqslant i \leqslant d < k$ *and* $n - k - s(k + d - 2i) \geqslant d - i$*. Then*

(i) $M(i;d,k,n)$ *is an* s^{e_2}*-disjunct matrix, where* $e_2 = \binom{k-s}{i-s}\binom{n-k-s(k+d-2i)}{d-i} - 1$;
(ii) *For a given* k*, if* $i < d$*, then* $\lim_{n\to\infty} \frac{e_2+1}{e_1+1} = \infty$.

Proof. (i) Let $B_0, B_1, \ldots, B_s \in \binom{[n]}{k}$ be any $s+1$ distinct columns of $M(i;d,k,n)$. Then, for each $j \in [s]$, there exists an x_j such that $x_j \in B_0 \setminus B_j$. Suppose $X_0 = \{x_j \mid 1 \leqslant j \leqslant s\}$. Then $X_0 \subseteq B_0$, and $X_0 \nsubseteq B_j$ for each $j \in [s]$. Note that the number of i-subsets of B_0 containing X_0 is $\binom{k-|X_0|}{i-|X_0|} = \binom{k-|X_0|}{k-i}$. Since $\binom{k-|X_0|}{k-i}$ is decreasing for $1 \leqslant |X_0| \leqslant s$ and gets its minimum at $|X_0| = s$, the number of i-subsets of B_0 containing X_0 is at least $\binom{k-s}{k-i}$.

Let A_0 be an i-subset of B_0 containing X_0. Then $|A_0 \cap B_j| < i$ for each $j \in [s]$. Let $D \in \binom{[n]}{d}$ satisfying $|D \cap B_0| = i$. If there exists $j \in [s]$ such that $|D \cap B_j| = i$, then $|B_0 \cap B_j| \geqslant |D \cap B_0 \cap B_j| \geqslant 2i - d$. Suppose $|B_0 \cap B_j| \geqslant 2i - d$ for each $j \in [s]$. Since $|\bigcup_{0 \leqslant j \leqslant s} B_j| \leqslant k + s(k + d - 2i)$, the number of d-subsets D of $[n]$ containing A_0 satisfying $|D \cap B_0| = i$ and $|D \cap B_j| \neq i$ for each $j \in [s]$ is at least $\binom{n-k-s(k+d-2i)}{d-i}$. Then the number of d-subsets D containing X_0 in $\binom{[n]}{d}$ satisfying $|D \cap B_0| = i$ and $|D \cap B_j| \neq i$ for each $j \in [s]$ is at least $\binom{k-s}{i-s}\binom{n-k-s(k+d-2i)}{d-i}$. Therefore, (i) holds.

(ii) is straightforward by (i) and Theorem 1. $\square$

Example 1. $M(5,7,50)$ is fully $1^{14}, 2^9$ and 3^5-disjunct, but $M(3;5,7,50)$ is $1^{9989}, 2^{2324}$ and 3^{299}-disjunct; $M(4,5,13)$ is fully 1^3 and 2^2-disjunct, but $M(3;4,5,13)$ is 1^{29} and 2^5-disjunct.

Concluding remarks

(i) For given integers $d < k$ the following limit holds: $\lim_{n\to\infty} \frac{\binom{n}{d}}{\binom{n}{k}} = 0$. This shows that the test-to-item of $M(i;d,k,n)$ is small enough when n is large enough. By Theorem 2, our pooling designs are better than Macula's designs when n is large enough.

(ii) It seems to be interesting to compute e such that $M(i;d,k,n)$ is fully s^e-disjunct.

(iii) In [12], Nan and the first author discussed the similar construction of s^e-disjunct matrices in a finite vector space, but the number e is not well expressed. By the method of this paper, e may be larger. We will study this problem in a separate paper.

(iv) For positive integers $1 \leqslant d < k < n$, let I be a nonempty proper subset of $\{0, 1, \ldots, d\}$, and let $M(I; d, k, n)$ be the binary matrix with rows indexed with $\binom{[n]}{d}$ and columns indexed with $\binom{[n]}{k}$ such that $M(A, B) = 1$ if and only if $|A \cap B| \in I$. How about the error-tolerant property of $M(I; d, k, n)$?

Acknowledgments

We would like thank the referees for their valuable suggestions. This research is partially supported by NSF of China, NCET-08-0052, Langfang Teachers' College (LSZB201005), and the Fundamental Research Funds for the Central Universities of China.

References

[1] Y. Bai, T. Huang, K. Wang, Error-correcting pooling designs associated with some distance-regular graphs, Discrete Appl. Math. 157 (2009) 3038–3045.
[2] Y. Cheng, D. Du, Efficient constructions of disjunct matrices with applications to DNA library screening, J. Comput. Biol. 14 (2007) 1208–1216.
[3] Y. Cheng, D. Du, New constructions of one- and two-stage pooling designs, J. Comput. Biol. 15 (2008) 195–205.
[4] D. Du, F.K. Hwang, Pooling Designs and Nonadaptive Group Testing, Important Tools for DNA Sequencing, Ser. Appl. Math., vol. 18, World Scientific Publishing Co., Pte. Ltd., Hackensack, NJ, 2006.
[5] A.G. D'yachkov, F.K. Hwang, A.J. Macula, P.A. Vilenkin, C. Weng, A construction of pooling designs with some happy surprises, J. Comput. Biol. 12 (2005) 1127–1134.
[6] A.G. D'yachkov, A.J. Macula, P.A. Vilenkin, Nonadaptive and trivial two-stage group testing with error-correcting d^e-disjunct inclusion matrices, in: Entropy, Search, Complexity, in: Bolyai Soc. Math. Stud., vol. 16, Springer, Berlin, 2007, pp. 71–83.
[7] H. Huang, Y. Huang, C. Weng, More on pooling spaces, Discrete Math. 308 (2008) 6330–6338.
[8] T. Huang, K. Wang, C. Weng, More pooling spaces associated with some finite geometries, European J. Combin. 29 (2008) 1483–1491.
[9] T. Huang, C. Weng, Pooling spaces and non-adaptive pooling designs, Discrete Math. 282 (2004) 163–169.
[10] A.J. Macula, A simple construction of d-disjunct matrices with certain constant weights, Discrete Math. 162 (1996) 311–312.
[11] A.J. Macula, Error-correcting non-adaptive group testing with d^e-disjunct matrices, Discrete Appl. Math. 80 (1997) 217–222.
[12] J. Nan, J. Guo, New error-correcting pooling designs associated with finite vector spaces, J. Comb. Optim. 20 (2010) 96–100.
[13] H. Ngo, D. Du, New constructions of non-adaptive and error-tolerance pooling designs, Discrete Math. 243 (2002) 167–170.