Optim Lett (2009) 3:297–302
DOI 10.1007/s11590-008-0109-2

ORIGINAL PAPER

On the upper bounds of the minimum number of rows of disjunct matrices

Yongxi Cheng · Ding-Zhu Du · Guohui Lin

Received: 20 September 2008 / Accepted: 23 October 2008 / Published online: 22 November 2008

Abstract A 0-1 matrix is d-*disjunct* if no column is covered by the union of any d other columns. A 0-1 matrix is $(d; z)$-*disjunct* if for any column C and any d other columns, there exist at least z rows such that each of them has value 1 at column C and value 0 at all the other d columns. Let $t(d, n)$ and $t(d, n; z)$ denote the minimum number of rows required by a d-disjunct matrix and a $(d; z)$-disjunct matrix with n columns, respectively. We give a very short proof for the currently best upper bound on $t(d, n)$. We also generalize our method to obtain a new upper bound on $t(d, n; z)$.

Keywords Disjunct matrices · Cover free families · Superimposed codes

1 Introduction

A 0-1 matrix is d-*disjunct* if no column is covered by the union of any d other columns, by union we mean the bitwise boolean sum of these d column vectors. In other words, a 0-1 matrix is called d-disjunct if for any column C and any d other columns, there

The work of Y. Cheng and G. Lin is supported by Natural Science and Engineering Research Council (NSERC) of Canada, and the Alberta Ingenuity Center for Machine Learning (AICML) at the University of Alberta.
The work of D.-Z. Du is partially supported by National Science Foundation under grant No.CCF0621829.

Y. Cheng (✉) · G. Lin
Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada
e-mail: yongxi@cs.ualberta.ca; chengyx@gmail.com

G. Lin
e-mail: ghlin@cs.ualberta.ca

D.-Z. Du
Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083, USA
e-mail: dzdu@utdallas.edu

exists at least one row such that the row has value 1 at column C and value 0 at all d other columns. The same structure is also called *cover free family* [9,10,15] in combinatorics, and *superimposed code* [6,8,12] in information theory. It is called a d-disjunct matrix in group testing [4,11,13]. A 0-1 matrix is $(d; z)$-*disjunct* [8,13] if for any column C and any d other columns, there exist at least z rows such that each of them has value 1 at column C and value 0 at all the other d columns. Thus, d-disjunct is $(d; 1)$-disjunct. Besides other applications, d-disjunct and $(d; z)$-disjunct matrices form the basis for error-free and error-tolerant nonadaptive group testing algorithms, respectively. These algorithms have applications in many practical areas such as DNA library screening [2–4,14] and multi-access communications [16], etc.

Let $t(d, n)$ denote the minimum number of rows required by a d-disjunct matrix with n columns. The bounds on $t(d, n)$ have been extensively studied in the fields of combinatorics, information theory, and group testing, using different terminologies. For lower bounds, $t(d, n) = \Omega(\frac{d^2 \log n}{\log d})$ [7,10,15] (throughout the paper log is of base 2 if no base is specified). In particular, D'yachkov and Rykov [7] proved that $t(d, n) \geq \frac{d^2}{2 \log d}(1 + o(1)) \log n$, which is the best lower bound so far. For upper bounds on $t(d, n)$, it is known that $t(d, n) = O(d^2 \log n)$ [8,11]. In [11] (also see [4, p. 57]), Hwang and Sós gave a greedy type construction which results in $t \times n$ d-disjunct matrices with $t \leq 16d^2(1 - \log_3 2 + (\log_3 2) \log_2 n)$. In [8], D'yachkov et al. obtained the following asymptotic upper bound on $t(d, n)$ with a rather involved proof, which is currently the best.

Theorem 1.1 (D'yachkov et al. [8]) *For d constant and $n \to \infty$, $t(d, n) \leq \frac{d}{A_d}[1 + o(1)] \log n$, where $A_d = \max_{0 \leq p \leq 1} \max_{0 \leq P \leq 1} \{-(1-P)\log(1-p^d) + d[P \log \frac{p}{P} + (1-P) \log \frac{1-p}{1-P}]\}$. Moreover, $A_d \to \frac{1}{d \log e}$ as $d \to \infty$.*

For $(d; z)$-disjunct matrices, let $t(d, n; z)$ denote the minimum number of rows required by a $(d; z)$-disjunct matrix with n columns. For given d and z, D'yachkov et al. [8] studied $\lim_{n \to \infty} \frac{\log n}{t}$ among others, and they proved that $t(d, n; z) \geq c[\frac{d^2 \log n}{\log d} + (z - 1)d]$ where c is a constant.

In this paper, by using the concept of q-*ary* $(d, 1)$-*disjunct* matrices [4,5] and the probabilistic method (see, e.g., [1]), we give a very short proof for the currently best upper bound on $t(d, n)$. In contrast to the previous result in [8] (Theorem 1.1) which is an asymptotic upper bound, our upper bound on $t(d, n)$ does not contain the asymptotic term $o(1)$. Also, we generalize our method to obtain a new upper bound on $t(d, n; z)$. Since our new proof is very short and concise, we hope that it can shed new light on this old problem and stimulate new research on it.

2 Upper bounds on $t(d, n)$

In this section we prove the following theorem.

Theorem 2.1 *For $n > d \geq 1$, $t(d, n) \leq \frac{d+1}{B_d} \log n$, where $B_d = \max_{q > 1} \frac{-\log[1-(1-\frac{1}{q})^d]}{q}$. Moreover, $B_d \to \frac{1}{d \log e}$ as $d \to \infty$.*

Before the proof, we first introduce the concept of *q-ary $(d, 1)$-disjunct* matrix. A matrix is called q-ary $(d, 1)$-disjunct if it is q-ary, and for any column C and any set D of d other columns, there exists an element in C such that the element does not appear in any column of D in the same row.

As described in [4,5], one can transform a q-ary $(d, 1)$-disjunct matrix M to a (binary) d-disjunct matrix M' as follows. Replace each row R_i of M by several rows indexed with entries of R_i. For each entry x of R_i, the row with index x is obtained from R_i by turning all x's into 1's and all others into 0's. From this transformation, we have the following theorem which is useful in our proof.

Theorem 2.2 (Theorem 3.6.1 in [4]) *A $t \times n$ q-ary $(d, 1)$-disjunct matrix M yields a $t' \times n$ d-disjunct matrix M' with $t' \le tq$.*

Now we are ready to prove Theorem 2.1.

Proof of Theorem 2.1 Given $n > d \ge 1$, first construct a random $t \times n$ q-ary ($q > 1$) matrix M with each entry assigned randomly and uniformly from $\{1, 2, \ldots, q\}$, where q and t will be specified later. For each column C and a set D of d other columns, for each element c_i ($i = 1, 2, \ldots, t$) of C, the probability that c_i appears in some column of D in the same row is $1 - (1 - \frac{1}{q})^d$. Thus the probability that every element of C appears in some column of D in the same row is $[1 - (1 - \frac{1}{q})^d]^t$. M is not $(d, 1)$-disjunct if and only if there exist a column C and a set D of d other columns such that the above holds. Therefore, the probability that M is not $(d, 1)$-disjunct is no more than $(n - d)\binom{n}{d}[1 - (1 - \frac{1}{q})^d]^t$.

We try to minimize tq, the number of rows of the d-disjunct matrix M' as in Theorem 2.2, under the condition that q and t satisfy

$$n^{d+1}\left[1 - \left(1 - \frac{1}{q}\right)^d\right]^t \le 1. \tag{2.1}$$

Notice that Eq. (2.1) implies $(n-d)\binom{n}{d}[1-(1-\frac{1}{q})^d]^t < 1$, thus the probability that M is $(d, 1)$-disjunct is greater than zero. Therefore, by probabilistic argument Eq. (2.1) implies the existence of a $t \times n$ q-ary $(d, 1)$-disjunct matrix, and so a d-disjunct matrix with n columns and at most tq rows.

To satisfy Eq. (2.1), let $t = \frac{(d+1)\log n}{-\log[1-(1-\frac{1}{q})^d]}$. Define $B_d(q) = \frac{-\log[1-(1-\frac{1}{q})^d]}{q}$, then $tq = \frac{(d+1)\log n}{B_d(q)}$. Let q_0 be the point that maximizes $B_d(q)$, and let $B_d = B_d(q_0)$ (one can estimate that $q_0 = \Theta(d)$ and $B_d = \Theta(\frac{1}{d})$, since the proof here can stand alone without this observation, we put it in appendix). By assigning $q = q_0$, we obtain

$$t(d, n) \le (tq)|_{q=q_0} = \frac{(d+1)\log n}{B_d(q_0)} = \frac{(d+1)\log n}{B_d}.$$

Finally, we estimate B_d as $d \to \infty$. Since $(1 - \frac{1}{q})^q < \frac{1}{e}$ for $q > 1$, $(1 - \frac{1}{q})^d < (\frac{1}{e})^{\frac{d}{q}} = e^{-\frac{d}{q}}$, and $-\log[1 - (1 - \frac{1}{q})^d] < -\log(1 - e^{-\frac{d}{q}})$. It follows that

$B_d(q) = \frac{-\log[1-(1-\frac{1}{q})^d]}{q} < \frac{-\log(1-e^{-\frac{d}{q}})}{q} = \frac{1}{d\ln 2}[-\frac{d}{q}\ln(1-e^{-\frac{d}{q}})]$. Let $x = e^{-\frac{d}{q}}$, then $-\frac{d}{q} = \ln x$, and $B_d(q) < \frac{1}{d\ln 2}\ln x\ln(1-x)$. Since $\ln x\ln(1-x)$ achieves its maximum at $x = \frac{1}{2}$, we obtain $B_d(q) < \frac{\ln 2}{d}$ for $q > 1$. Thus $B_d < \frac{\ln 2}{d}$ for $d \geq 1$. On the other hand, when q satisfies $(1-\frac{1}{q})^d = \frac{1}{2}$, as $d \to \infty$, it is easy to see that $\frac{q}{d} \to \frac{1}{\ln 2}$, and $B_d(q) = \frac{1}{q} \to \frac{\ln 2}{d}$. Therefore, as $d \to \infty$, $B_d \to \frac{\ln 2}{d} = \frac{1}{d\log e}$. □

3 New upper bounds on $t(d, n; z)$

In this section, we generalize the above method to obtain new upper bounds for $(d; z)$-disjunct matrices. We establish the following theorem.

Theorem 3.1 *For d, z constants, and $n \to \infty$, $t(d, n; z) \leq \frac{d+1}{B_d}\log n + \frac{z}{B_d}\log\log n + O(1)$, where $B_d = \max_{q>1}\frac{-\log[1-(1-\frac{1}{q})^d]}{q}$. Moreover, $B_d \to \frac{1}{d\log e}$ as $d \to \infty$.*

A q-ary matrix is called $(d, 1; z)$-*disjunct* if for any column C and any set D of d other columns, there exists at least z elements in C such that each of these elements does not appear in any column of D in the same row. Clearly, by using the same method mentioned above, one can transform a $t \times n$ q-ary $(d, 1; z)$-disjunct matrix to a $(d; z)$-disjunct matrix with n columns and at most tq rows.

Proof of Theorem 3.1 For given n, d and z, similarly we construct a random $t \times n$ q-ary $(q > 1)$ matrix M with each entry assigned randomly and uniformly from $\{1, 2, \ldots, q\}$, q and t will be specified later. For each column C and a set D of d other columns, for each element c_i of C, the probability that c_i appears in some column of D in the same row is $1-(1-\frac{1}{q})^d$. Thus the probability that there exist $t-z+1$ elements of C such that each of them appears in some column of D in the same row is at most $\binom{t}{t-z+1}[1-(1-\frac{1}{q})^d]^{t-z+1} = \binom{t}{z-1}[1-(1-\frac{1}{q})^d]^{t-z+1}$. M is not $(d, 1; z)$-disjunct if and only if there exists a column C and a set D of d other columns such that the above holds. Therefore, the probability that M is not $(d, 1; z)$-disjunct is no more than $(n-d)\binom{n}{d}\binom{t}{z-1}[1-(1-\frac{1}{q})^d]^{t-z+1}$.

We want to minimize tq, the number of rows of the corresponding $(d; z)$-disjunct matrix, under the condition that

$$n^{d+1}t^z\left[1-\left(1-\frac{1}{q}\right)^d\right]^{t-z} \leq 1. \tag{3.1}$$

Notice that Eq. (3.1) implies $(n-d)\binom{n}{d}\binom{t}{z-1}[1-(1-\frac{1}{q})^d]^{t-z+1} < 1$. Thus the probability that M is $(d, 1; z)$-disjunct is greater than zero, which similarly implies the existence of a $t \times n$ q-ary $(d, 1; z)$-disjunct matrix, and a $(d; z)$-disjunct matrix with n columns and at most tq rows.

Let q_0 be the point that maximizes $B_d(q) = \frac{-\log[1-(1-\frac{1}{q})^d]}{q}$. Assign $q = q_0$. To satisfy Eq. (3.1), which is equivalent to $(d+1)\log n + z\log t \leq -(t-z)\log[1-(1-\frac{1}{q_0})^d]$, let $t = \frac{(d+1)\log n}{-\log[1-(1-\frac{1}{q_0})^d]} + z + t_1$. Then, t_1 should satisfy

$$z \log \left\{ \frac{(d+1)\log n}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} + z + t_1 \right\} \le -t_1 \log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]. \tag{3.2}$$

Let $t_1 = \frac{z \log\log n}{-\log[1-(1-\frac{1}{q_0})^d]} + t_2$, from Eq. (3.2), t_2 should satisfy that

$$\begin{aligned} &\frac{z}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} \log \left\{ \frac{(d+1)}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} \right. \\ &\left. + \frac{1}{\log n}\left(\frac{z\log\log n}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} + z + t_2 \right) \right\} \le t_2. \end{aligned} \tag{3.3}$$

For d and z constants (thus q_0 is also constant), as $n \to \infty$, the minimum value of t_2 satisfying Eq. (3.3) is $t_2 = \frac{z}{-\log[1-(1-\frac{1}{q_0})^d]} \log \frac{(d+1)}{-\log[1-(1-\frac{1}{q_0})^d]} = O(1)$. Thus,

$$t = \frac{(d+1)\log n}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} + \frac{z\log\log n}{-\log\left[1-\left(1-\frac{1}{q_0}\right)^d\right]} + O(1)$$

satisfies Eq. (3.1) (where the constant term z in t is absorbed in O(1)). Therefore, the number of rows of the corresponding $(d; z)$-disjunct matrix is at most

$$tq_0 = \frac{d+1}{B_d}\log n + \frac{z}{B_d}\log\log n + O(1),$$

where $B_d = B_d(q_0) = \max_{q>1} \frac{-\log[1-(1-\frac{1}{q})^d]}{q}$. Also, $B_d \to \frac{1}{d\log e}$ as $d \to \infty$, as proved in Theorem 2.1. □

Appendix A: Estimating q_0 and B_d

Lemma A.1 *Given $d \ge 1$, let $q_0 = q_0(d)$ be the point that maximizes $B_d(q) = \frac{-\log[1-(1-\frac{1}{q})^d]}{q}$ for $q > 1$. Then, as $d \to \infty$, $q_0(d) = \Theta(d)$, and $B_d = B_d(q_0) = \Theta(\frac{1}{d})$.*

Proof Notice that if q_1 satisfies $(1-\frac{1}{q_1})^d = \frac{1}{2}$, then $q_1 = \Theta(d)$ since $\frac{q_1}{d} \to \frac{1}{\ln 2}$ as $d \to \infty$. Moreover, $B_d(q_1) = \frac{1}{q_1} = \Theta(\frac{1}{d})$. We prove the lemma by contradiction. First assume that $q_0 = O(d)$ does not hold, that is, for any $c > 0$ and any $d_0 > 0$, there exists $d > d_0$ such that $q_0(d) > cd$. Then, since $\frac{q_0}{d} > c$, as $c \to \infty$, $B_d(q_0)d = \frac{-\log[1-(1-\frac{1}{q_0})^d]}{q_0} d \sim \frac{-\log[1-(1-\frac{d}{q_0})]}{q_0} d = \frac{\log\frac{q_0}{d}}{\frac{q_0}{d}} = o(1)$ (here by $a \sim b$ we mean that $\lim_{c\to\infty}\frac{a}{b} = 1$). Thus, $B_d(q_0) = \frac{o(1)}{d}$. It contradicts since q_0 is the maximum point of $B_d(q)$ and $B_d(q_1) = \Theta(\frac{1}{d})$ with $(1-\frac{1}{q_1})^d = \frac{1}{2}$. On the other hand, assume that $q_0 = \Omega(d)$ does not hold, that is, for any $c > 0$ and any $d_0 > 0$, there exists $d > d_0$ such that $q_0(d) < cd$. Then, $B_d(q_0)d = \frac{-\log[1-(1-\frac{1}{q_0})^d]}{q_0} d = \frac{-\ln\{1-[(1-\frac{1}{q_0})^{q_0}]^{\frac{d}{q_0}}\}}{q_0 \ln 2} d$. Since $0 < (1-\frac{1}{q_0})^{q_0} < \frac{1}{e}$ for $q_0 > 1$, as $c \to 0$, $\frac{d}{q_0} > \frac{1}{c} \to \infty$, and $[(1-\frac{1}{q_0})^{q_0}]^{\frac{d}{q_0}} < e^{-\frac{d}{q_0}} \to 0$. Thus $B_d(q_0)d \sim \frac{[(1-\frac{1}{q_0})^{q_0}]^{\frac{d}{q_0}}}{q_0 \ln 2} d = \frac{1}{\ln 2}\frac{d}{q_0}[(1-\frac{1}{q_0})^{q_0}]^{\frac{d}{q_0}} < \frac{1}{\ln 2}\frac{d}{q_0}e^{-\frac{d}{q_0}} = o(1)$, which also contradicts (here by $a \sim b$ we mean that $\lim_{c\to 0}\frac{a}{b} = 1$). Therefore, $q_0(d) = \Theta(d)$. Then, $(1-\frac{1}{q_0})^d < 1$ is $\Theta(1)$, and thus $B_d(q_0) = \frac{\Theta(1)}{q_0} = \Theta(\frac{1}{d})$. □

References

1. Alon, N., Spencer, J.H.: The Probabilistic Method. Wiley, New York (1992)
2. Balding, D.J., Bruno, W.J., Knill, E., Torney, D.C.: A comparative survey of non-adaptive pooling designs. In: Genetic Mapping and DNA Sequencing, pp. 133–154. Springer, New York (1996)
3. Bruno, W.J., Balding, D.J., Knill, E., Bruce, D.C., Doggett, N.A., Sawhill, W.W., Stallings, R.L., Whittaker, C.C., Torney, D.C.: Efficient pooling designs for library screening. Genomics **26**, 21–30 (1995)
4. Du, D.Z., Hwang, F.K.: Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing. World Scientific, Singapore (2006)
5. Du, D.Z., Hwang, F.K., Wu, W., Znati, T.: New construction for transversal design. J. Comput. Biol. **13**, 990–995 (2006)
6. D'yachkov, A.G., Macula, A.J., Rykov, V.V.: New constructions of superimposed codes. IEEE Trans. Inform. Theory **46**, 284–290 (2000)
7. D'yachkov, A.G., Rykov, V.V.: Bounds of the length of disjunct codes. Problems Control Inform. Theory **11**, 7–13 (1982)
8. D'yachkov, A.G., Rykov, V.V., Rashad, A.M.: Superimposed distance codes. Problems Control Inform. Theory **18**, 237–250 (1989)
9. Erdös, P., Frankl, P., Füredi, Z.: Families of finite sets in which no set is covered by the union of r others. Israel J. Math. **51**, 79–89 (1985)
10. Füredi, Z.: On r-cover-free families. J. Comb. Theory Ser. A **73**, 172–173 (1996)
11. Hwang, F.K., Sós, V.T.: Non-adaptive hypergeometric group testing. Studia Scient. Math. Hungarica. **22**, 257–263 (1987)
12. Kautz, W.H., Singleton, R.C.: Nonrandom binary superimposed codes. IEEE Trans. Inform. Theory **10**, 363–377 (1964)
13. Macula, A.J.: Error-correcting nonadaptive group testing with d^e-disjunct matrices. Discrete Appl. Math. **80**, 217–222 (1997)
14. Ngo, H.Q., Du, D.Z.: A survey on combinatorial group testing algorithms with applications to DNA library screening. DIMACS: Series in Discrete Mathematics and Theoretical Computer Science, vol. 55, pp. 171–182. American Mathematical Society, Providence (2000)
15. Ruszinkó, M.: On the upper bound of the size of the r-cover-free families. J. Comb. Theory Ser. A **66**, 302–310 (1994)
16. Wolf, J.K.: Born again group testing: multiaccess communications. IEEE Trans. Inform. Theory **IT-31**, 185–191 (1998)