Group Testing with Multiple Mutually-Obscuring Positives

Hong-Bin Chen[1] and Hung-Lin Fu[2]

[1] Institute of Mathematics, Academia Sinica, Taipei 10617, Taiwan
hbchen@math.sinica.edu.tw
[2] Department of Applied Mathematics, National Chiao Tung University, Hsinchu 30050, Taiwan
hlfu@math.nctu.edu.tw

In memory of a great mathematician and information scientist Rudolf Ahlswede

Abstract. Group testing is a frequently used tool to identify an unknown set of defective (positive) elements out of a large collection of elements by testing subsets (pools) for the presence of defectives. Various models have been studied in the literature. The most studied case concerns only two types (defective and non-defective) of elements in the given collection. This paper studies a novel and natural generalization of group testing, where more than one type of defectives are allowed with an additional assumption that certain obscuring phenomena occur among different types of defectives. This paper proposes some algorithms for this problem, trying to optimize different measures of performance: the total number of tests required, the number of stages needed to perform all tests and the decoding complexity.

Keywords: pooling design, group testing, selectors.

1 Introduction

The classical group testing problem is described as follows: Given a set $\mathcal{N}$ of n items consisting of two types of items, a set $\mathcal{P}$ of positive items with $|\mathcal{P}| \leq d$ and the others being negative items, the goal is to identify $\mathcal{P}$ in an efficient manner by using group tests. A test (pool) can be applied to any subset of items in $\mathcal{N}$ with two possible outcomes: a negative outcome indicates that there is no positive in the test while a positive outcome indicates that at least one positive is in the test. The concept of group testing originated from the application of blood testing during World War II. Afterwards, it has been also found applications in molecular biology, including screening clone libraries [3], sequencing by hybridization [31], yeast one-hybrid screens [36], and recently, the mapping of protein-protein interactions [37]. Additionally, group testing has proved relevant in other fields such as multiple access communication [4], image compression [27] and more recently data gathering in sensor networks [28]. For general references, readers may refer to the books [18,19].

H. Aydinian, F. Cicalese, and C. Deppe (Eds.): Ahlswede Festschrift, LNCS 7777, pp. 557–568, 2013.

Due to a diversity of its applications, there has been many models that were proposed and studied in the literature. For example, the inhibitor model [5,6,8,26] where the presence of an inhibitor can somehow cancel the effect of positive elements, the complex model [1,2,10,12,35] where positive reactions are caused by certain sets of elements rather than a single one of elements, the threshold model [9,15,11,17] where two thresholds are given for the conditions of positive reactions and negative reactions to occur, the interference model [7,16,20] where two or more positive elements appearing in a pool can interfere with each other so that the positive reaction cannot be detected, and more others.

In this paper we study a generalization of group testing as follows. Consider a set $\mathcal{N}$ of n items which is known to contain s types of positive items $P_1, P_2, \cdots, P_s$, where $|P_i| \leq p_i$, and the others being negative items. The task is to classify all items in $\mathcal{N}$ with as few tests as possible. For a test $Q \subseteq \mathcal{N}$, define

$$I_Q \equiv \{i : Q \cap P_i \neq \varnothing\}.$$

Then the outcome of a test Q will be given according to the following rules:

- If $I_Q = \{i\}$, then the response will be "i-positive".
- If $I_Q = \varnothing$, then the response will be "negative".
- If $|I_Q| \geq 2$, then the response can be either "negative" or be "i-positive" for some $i \in I_Q$ (not knowing which i).

For example, given a test Q with $I_Q = \{1, 3, 4\}$, the outcome of the test Q can be any and exactly (but not knowing which) one of the four cases: negative, 1-positive, 3-positive and 4-positive. We refer to this problem as the *Multiple Mutually-Obscuring Positives* (MMOP) problem. Obviously, for the case $s = 1$, the MMOP problem is exactly the same as the classical group testing. For the case $s = 2$, it is coincident to the coin-weighing problem with test-type device (such as a spring balancer or electronic scale) where, given a set of coins and some of them are counterfeit (too heavy or too light), the task is not only to identify all counterfeit coins but also to make them classified as heavy or light. Notably, the MMOP problem is not a generalization of the "mutually-obscuring problem" discussed in [7,16,20].

In this paper, we provide a unified technique to deal with the MMOP problem for general s. In the next section, we propose an efficient *nonadaptive algorithm*, i.e., all tests are set up in advance and thus can be performed simultaneously without any information of outcomes of other tests. In particular, the proposed algorithm can be decoded in polynomial time. Section 3 provides a 2-stage algorithm for this problem. Instrumental to the result is based on a combinatorial structure, (k, m, n)-*selector*, first introduced by De Bonis, Gąsieniec and Vaccaro [6] in the context of designing efficient trivial 2-stage pooling strategies on the classic pooling design problem. We propose a new point of view for the mentioned selectors. This enables us to construct the combinatorial tool easily. Probabilistic constructions are provided and numerical results show that our constructions are slightly better than the currently best known result by De Bonis et al. [6].

2 Nonadaptive Algorithms

In order to present our results, we now introduce some notations and definitions. Note that n and p_i's are given in advance and we assume $\sum_{i=1}^{s} p_i \ll n$. Throughout this paper, a pooling design is represented by a 0-1 matrix M where columns are the set of objects, rows are the set of tests, and cell $(i,j) = 1$ signifies that the j-th object is in the i-th test and $(i,j) = 0$ for otherwise. For convenience, a column (row) can be treated as the set of row (column) indices where the column (row) has a 1, respectively. For any two columns C and C', we denote $C \cup C'$ as the boolean sum of C and C'. We say that a set X of columns appears (or is contained) in a row if all columns in X have a 1-entry in the row. A pool is called an i-positive pool if its outcome is i-positive, and a non-i-positive pool if it is not i-positive. For a column C, denote by $t_i(C)$ the number of i-positive pools in which column C appears. Likewise, denote by $t_{\bar{i}}(C)$ the number of non-i-positive pools in which column C appears.

Consider a fixed family $T = \{T_1, T_2, \cdots, T_t\}$ with $T_i \subseteq \mathcal{N}, 1 \le i \le t$. Let R_i, $1 \le i \le s$, be arbitrary disjoint subsets of $\mathcal{N}$ and let $J = \cup_{i=1}^{s} R_i$. We define the *syndrome vector* of J in T by $\phi_T(J) = (\phi_1(J), \phi_2(J), \cdots, \phi_t(J))$, where

$$\phi_j(J) = \{i : T_j \cap R_i \neq \varnothing\}.$$

For any two distinct J_0 and J_1, we say their syndromes $\phi_T(J_0)$ and $\phi_T(J_1)$ are *different*, denoted by $\phi_T(J_0) \not\sim \phi_T(J_1)$, if and only if there exists some $j \in [t]$ and $i \in \{0,1\}$ such that $\phi_j(J_{1-i}) = \{k\}$ for some $k \in [s]$ and $k \notin \phi_j(J_i)$. Denote by $\phi_T(J_0) \sim \phi_T(J_1)$ if they are *coincident* (not different).

Definition 1. *Let $T = \{T_1, T_2, \cdots, T_t\}$ with $T_i \subseteq \mathcal{N}, 1 \le i \le t$. We say the family T is MMOP-separable if*

$$\phi_T(J_0) \not\sim \phi_T(J_1)$$

for any two distinct $J_0, J_1 \subseteq \mathcal{N}$ with $J_0 = \cup_{i=1}^{s} R_i^0$ and $J_1 = \cup_{i=1}^{s} R_i^1$, where R_i^ℓ's, $\ell \in \{0,1\}$, are disjoint subsets of $\mathcal{N}$ and $|R_i^\ell| \le p_i$ for $1 \le i \le s$.

Lemma 1. *MMOP-separability is a sufficient and necessary condition for the MMOP problem.*

Proof. The lemma follows by definition immediately. □

We first present a lower bound on the number of tests required for any nonadaptive algorithms for the MMOP model. This lower bound is obtained by establishing a connection to disjunct matrices (equivalently, superimposed codes or cover-free families).

Definition 2. *[30] A binary matrix is called d-disjunct if for any $d+1$ columns $C_0, C_1, \cdots, C_d$,*

$$\left| C_0 \setminus \bigcup_{i=1}^{d} C_i \right| \ge 1.$$

It is well-known [21] that disjunct matrices of size $t \times n$ have a lower bound $t = \Omega(d^2 \log n / \log d)$ and an upper bound $t = O(d^2 \log n)$. The literature contains many studies (see [19] and references therein) on the construction of disjunct matrices (sometimes called superimposed codes and cover-free families). One of the most common approaches is to control the number of intersections of any two columns to guarantee the disjunctness property. For recognition, we refer disjunct matrices with this particular structure as "w/λ-disjunct".

Definition 3. *[30] A binary matrix is w/λ-disjunct if the following properties both hold: (1) every column has more than w 1-entries; (2) any two distinct columns intersect at no more than λ rows.*

It is easy to see that an w/λ-disjunct matrix is $\lfloor \frac{w}{\lambda} \rfloor$-disjunct and so has the lower bound $\Omega(d^2 \log n / \log d)$ on the number of rows where $d = \lfloor \frac{w}{\lambda} \rfloor$. Previous results [21,22,29,13,14,32] have also shown that this particular structure can achieve the best known upper bound $O(d^2 \log n)$ on the number of rows for d-disjunct matrices.

For short, let $d = \sum_{i=1}^{s} p_i$ in the rest of the paper. Note that asymptotic results presented in the paper are under the assumption that d is constant and n approaches to infinity.

Theorem 1. *Let $\mathcal{N}$ be a set of n items which is known to contain s types of positive items $P_1, P_2, \cdots, P_s$, where $|P_i| \leq p_i$, and the others being negative items. Then any nonadaptive algorithm for the MMOP problem requires $\Omega(d^2 \log n / \log d)$ tests to classify all positive items.*

Proof. To prove this theorem, it suffices to show that MMOP-separable implies $(d-1)$-disjunct. Then, by Lemma 1 and the well-known lower bound for disjunct matrices, we get the desired bound. Suppose to the contrary that there exists a family $T = \{T_1, T_2, \cdots, T_t\}$ of subset of $\mathcal{N}$ and its corresponding matrix M is not a $(d-1)$-disjunct matrix of size $t \times n$. Then there exist d columns $C_0, C_1, \cdots, C_{d-1}$ in M such that $|C_0 \setminus \bigcup_{j=1}^{d-1} C_i| = 0$. That means for every row T_i where C_0 appears there exists some $j \in \{1, 2, \cdots, d-1\}$ such that C_j also appears in the row T_i. Consider the two subsets $J_0 = \{C_1, \cdots, C_{d-1}\}$ and $J_1 = \{C_0, C_1, \cdots, C_{d-1}\}$. Then, obviously, J_0 and J_1 are distinct but their syndromes $\phi_T(J_0)$ and $\phi_T(J_1)$ are not different, i.e., there does not exist $j \in [t]$ and $i \in \{0, 1\}$ such that $\phi_j(J_{1-i}) = \{k\}$ for some $k \in [s]$ and $k \notin \phi_j(J_i)$. Thus, by Lemma 1, the $(d-1)$-disjunctness is a necessary condition for any nonadaptive strategy and the bound follows immediately. □

Next, we propose a nonadaptive algorithm for the considered problem. In particular, our algorithm can be decoded in polynomial time to recover all positives from outcomes.

Theorem 2. *Let $\mathcal{N}$ be a set of n items which is known to contain s types of positive items $P_1, P_2, \cdots, P_s$, where $|P_i| \leq p_i$, and the others being negative items. A (w/λ)-disjunct matrix of n columns with $w/\lambda > d$ can solve the MMOP problem using $O(d^2 \log n)$ tests and $O(d^2 n \log n)$ time.*

Proof. Let M be a (w/λ)-disjunct matrix of n columns with $w/\lambda > d$ and use it as the pooling design. Consider an arbitrary P_i for some i and let R_i be an element in P_i. Observe that R_i appears in a non-i-positive pool only when the pool contains another item $R_j \in P_j$ for some $j \neq i$. Since M is a (w/λ)-disjunct matrix, we have that $t_{\bar{i}}(R_i) \leq \lambda \sum_{k \neq i} p_k$. For an item $C \notin P_i$, C appears in an i-positive pool only when the pool contains some items of P_i. Hence, we conclude that $t_{\bar{i}}(C) \geq w - \lambda p_i$ since M is (w/λ)-disjunct.

By the above discussion along with the condition $w/\lambda > d$, we have that

$$t_{\bar{i}}(R_i) \leq \lambda \sum_{k \neq i} p_k < w - \lambda p_i \leq t_{\bar{i}}(C)$$

for any $R_i \in P_i$ and $C \notin P_i$. Thus, we can separate all items in P_i from those not in P_i through counting $t_{\bar{i}}(C)$ for each $C \in \mathcal{N}$. Since P_i is chosen arbitrarily, all items in $\mathcal{N}$ can be classified in a similar way.

For the decoding issue, our algorithm only needs to compute $t_{\bar{i}}(C)$ for each item C in $\mathcal{N}$. This can be done easily by going through each entry in the column C once. Hence, the decoding complexity is at most $O(d^2 n \log n)$ time. □

The following is the pseudo code of our decoding algorithm.

Algorithm 1 CLASSIFICATION

```
Use a (w/λ)-disjunct matrix with w/λ > Σ_{i=1}^{s} p_i as a pooling design.
P_i ← ∅, for 1 ≤ i ≤ s.
for each item C ∈ 𝒩 do
    if t_ī(C) ≤ λ Σ_{j≠i} p_j for some i then
        P_i ← P_i ∪ {C}
Return P_i for 1 ≤ i ≤ s.
```

3 A 2-Stage Algorithm

Theorem 1 shows that any nonadaptive algorithm for the MMOP problem requires $\Omega(d^2 \log n / \log d)$ tests. However, the information-theoretic lower bound for algorithms without any constraint on the number of stages reduces down to $\log_{s+1} \binom{n}{d} = \Theta(d \log(n/d) / \log(s+1))$. This section shall provide a 2-stage algorithm for the MMOP problem that uses $O(d \log(n/d))$ tests.

Definition 4. *[6] Given integers k, m and n with $1 \leq m \leq k \leq n$, we say that a binary $t \times n$ matrix M is a (k, m, n)-selector if any submatrix of M subject to k out of n arbitrary columns contains at least m distinct rows of the identity matrix I_k.*

The integer t is called the size of the (k, m, n)-selector. Denote by $t_s(k, m, n)$ the minimum size of a (k, m, n)-selector. De Bonis, Gąsieniec and Vaccaro [6] suggested a way to construct a (k, m, n)-selector by searching for a vertex cover on a properly defined hypergraph and derived the following result, which is the best known asymptotical bound.

Theorem 3. *[6] For any integers k, m, n with $1 \leq m \leq k < n$,*

$$t_s(k, m, n) < \frac{ek^2}{k - m + 1} \ln(n/k) + \frac{ek(2k - 1)}{k - m + 1},$$

where e is the base of natural logarithm.

By definition, we have the following result immediately.

Lemma 2. *A (k, m, n)-selector is a $(k - i, m - i, n)$-selector for any integer i with $1 \leq i \leq m - 1$.*

Theorem 4. *For all integers n and k with $n \geq k \geq 2d$, there exists a two-stage group testing algorithm for the MMOP problem that classifies all types of positives, and uses at most $t_s(k, 2d - 1, n) + k - d$ tests.*

Proof. Our algorithm works as follows. In the first stage, we test the pools associated with rows of a $(k, 2d - 1, n)$-selector M satisfying $n \geq k \geq 2d$ to find a set D' of at most $k-d$ suspicious candidates. In the second stage, all suspicious candidates are tested individually and simultaneously so as to classify all types of positives. To complete the proof, it suffices to show that after the first stage we can determine such a set D' that contains all positives with $|D'| \leq k - d$.

Let D be the set of actual positive items and let J be the set of the column indices associated with the set D. Suppose that there are exactly y pairwise distinct sets $J_1, J_2, \cdots, J_y$ of suspicious candidates and each of whose syndrome vectors agrees with the set D of all positives. Notice that $|J_i| \leq d$ for all i, and obviously $J \in \{J_1, J_2, \cdots, J_y\}$. If $|\bigcup_{i=1}^{y} J_i| \leq k - d$ as desired, then we are done. If $|\bigcup_{i=1}^{y} J_i| \geq k - d + 1$, then there must exist an integer ℓ with $2 \leq \ell \leq y$ satisfying

$$k - d + 1 \leq |\bigcup_{i=1}^{\ell} J_i| \leq k$$

because of $|J_i| \leq d$ and $k \geq 2d$. Let $|\bigcup_{i=1}^{\ell} J_i| = q$. Since $2d - 1 - (k - q) \geq d$, by Lemma 2, M is also a (q, d, n)-selector. Subject to the q columns associated with the set $\bigcup_{i=1}^{\ell} J_i$, the submatrix of M contains at least d distinct rows of the identity matrix I_q. That means at least d suspicious candidates that appear separately and individually in some rows without any other suspicious candidates of $\bigcup_{i=1}^{\ell} J_i$. For the coincidence of the syndromes of $J_1, J_2, \cdots, J_\ell$, each of these d or more suspicious candidates must belong to every J_i for $1 \leq i \leq \ell$, a contradiction to the assumption that J_i's are pairwise distinct sets with $|J_i| \leq d$ for each i. Hence, we can determine a set $D' = \bigcup_{i=1}^{y} J_i$ of cardinality at most $k - d$, as desired. □

By Theorem 4 with $k = 4d - 2$ and Theorem 3, we have the following result.

Corollary 1. *There exists a two-stage group testing algorithm for the MMOP problem that classifies all types of positives, and uses at most* $O(d\log(n/d))$ *tests.*

Note that for the particular case $s = 1$ (indeed the classic group testing problem) this result can be found in [22,23,24,25].

3.1 Improved Upper Bounds on the Size of (k, m, n)-Selectors

In this subsection, we exploit three different probabilistic methods to derive upper bounds on the size of (k, m, n)-selectors and present numerical results of these bounds and that in [6] for comparison. The three probabilistic methods can also be found in several studies [14,12,21,22,33,34] on different combinatorial structures.

Definition 5. *A matrix* $[m_{ij}]$ *is strongly* (d, r)*-disjunct if for any two disjoint sets* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *of columns with* $|\mathcal{C}_1| = d$ *and* $|\mathcal{C}_2| = r$*, there exists a row* k *and a column* $c \in \mathcal{C}_2$ *such that* $m_{kc} = 1$ *and* $m_{kj} = 0$ *for all* $j \in \mathcal{C}_1 \cup \mathcal{C}_2 \setminus \{c\}$.

Theorem 5. *A* (k, m, n)*-selector is the same as a strongly* $(m-1, k-m+1)$*-disjunct matrix of* n *columns.*

Proof. Suppose that M is a (k, m, n)-selector. For any two disjoint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ of columns with $|\mathcal{C}_1| = m - 1$ and $|\mathcal{C}_2| = k - m + 1$, the submatrix $M_{\mathcal{C}_1\cup\mathcal{C}_2}$ subject to $\mathcal{C}_1 \cup \mathcal{C}_2$ must contain m distinct rows of an identity matrix I_k, by the definition of (k, m, n)-selectors. Therefore, there exists at least one row i and a column $c \in \mathcal{C}_2$ (by pigeonhole principle) such that $m_{ic} = 1$ and $m_{ij} = 0$ for all $j \in \mathcal{C}_1 \cup \mathcal{C}_2 \setminus \{c\}$, as desired.

Suppose that M is a strongly $(m-1, k-m+1)$-disjunct matrix of n columns. For any fixed set $\mathcal{K}$ of k columns, we want to find m distinct rows of the identity matrix I_k in $M_{\mathcal{K}}$. Partition arbitrarily $\mathcal{K}$ into two disjoint sets $\mathcal{C}_1 = \{c_{1,1}, c_{1,2}, \cdots, c_{1,m-1}\}$ and $\mathcal{C}_2 = \mathcal{K} \setminus \mathcal{C}_1$ with $|\mathcal{C}_1| = m - 1$ and $|\mathcal{C}_2| = k - m + 1$, then we can find a row i_1 and a column $c_{2,1} \in \mathcal{C}_2$, such that $m_{i_1c_{2,1}} = 1$ and $m_{i_1j} = 0$ for all $j \in \mathcal{C}_1\cup\mathcal{C}_2\setminus\{c_{2,1}\}$ since M is strongly $(m-1, k-m+1)$-disjunct. Exchanging the column $c_{1,1}$ with $c_{2,1}$ such that $\mathcal{C}_1 = \{c_{2,1}, c_{1,2}, \cdots, c_{1,m-1}\}$ and $\mathcal{C}_2 = \mathcal{K} \setminus \mathcal{C}_1$, similarly we can find another row i_2 and a column $c_{2,2} \in \mathcal{C}_2$ satisfying the desired property, as in row i_1. Notice that the column $c_{2,2}$ chosen from $\mathcal{C}_2$ must be different from the column $c_{2,1}$ which is in the updated $\mathcal{C}_1$. Keep doing this process until $c_{1,j}$'s are all removed to $\mathcal{C}_2$, we then obtain a set $\{i_1, i_2, \cdots, i_m\}$ of m distinct rows of the identity matrix I_k in $M_{\mathcal{K}}$. Hence M is a (k, m, n)-selector. □

Next, we derive upper bounds by probabilistic methods on the minimum number of rows of strongly (d, r)-disjunct matrices, and consequently on the minimum size of selectors. Construct a $t_1 \times 2n$ $0-1$ matrix M where each entry is defined to be 1 with probability p and 0 with probability $1 - p$. We say that a column c_j is unsatisfied if there exists two disjoint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ of columns with $c_j \in \mathcal{C}_1 \cup \mathcal{C}_2$, $|\mathcal{C}_1| = d$ and $|\mathcal{C}_2| = r$ such that Definition 5 is not true for

any row. Then, given a fixed column c_j, the probability of the event that c_j is unsatisfied is

$$P(c_j \text{ is unsatisfied}) = \binom{2n-1}{d+r-1}\binom{d+r}{d}\left[1 - rp(1-p)^{d+r-1}\right]^{t_1}. \quad (1)$$

By simple technique of calculus, we take $p = \frac{1}{d+r}$ to minimize Equation (1) and obtain

$$P(c_j \text{ is unsatisfied}) = \binom{2n-1}{d+r-1}\binom{d+r}{d}\left[1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right]^{t_1}. \quad (2)$$

Setting

$$t_1 = \frac{\ln\left(\binom{2n-1}{d+r-1}\binom{d+r}{d}\right)}{-\ln\left(1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right)} + \ln 2, \quad (3)$$

the right-hand side of (2) is less than $1/2$, which implies that the expected value of the total number of unsatisfied columns in the matrix M does not exceed n. Hence, there must exist a matrix which is strongly (d, r)-disjunct and of size $t_1 \times n$.

Secondly, we construct a random $t_2 \times n$ $0-1$ matrix M where each entry is defined to be 1 with probability $\frac{1}{d+r}$ and 0 with probability $1 - \frac{1}{d+r}$. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two disjoint sets of columns with $|\mathcal{C}_1| = d$ and $|\mathcal{C}_2| = r$. Similarly, we say that the pair $(\mathcal{C}_1,\mathcal{C}_2)$ is unsatisfied if this pair does not satisfy the requirement of Definition 5. Then the probability of the event that $(\mathcal{C}_1,\mathcal{C}_2)$ is unsatisfied is

$$P((\mathcal{C}_1,\mathcal{C}_2) \text{ is unsatisfied}) = \left[1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right]^{t_2}. \quad (4)$$

Hence the expected value of the total number of unsatisfied pairs in the matrix M is

$$E[\text{unsatisfied pairs}] = \binom{n}{d+r}\binom{d+r}{d}\left[1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right]^{t_2}. \quad (5)$$

Setting

$$t_2 = \frac{\ln\left(\binom{n}{d+r}\binom{d+r}{d}\right)}{-\ln\left(1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right)}, \quad (6)$$

the right-hand side of Equation (5) is less than 1, which implies that the probability of the existence of a strongly (d, r)-disjunct matrix of size $t_2 \times n$ is greater than 0. Thus, there exists a strongly (d, r)-disjunct matrix of size $t_2 \times n$.

Remark: The above argument can be further extended to a high probability version. Given $0 < \epsilon < 1$, if we require $t_2 = \frac{\ln\left(\binom{n}{d+r}\binom{d+r}{d}\right) - \ln\epsilon}{-\ln\left(1 - \frac{r}{d+r}\cdot\left(\frac{d+r-1}{d+r}\right)^{d+r-1}\right)}$, then Equation (5) is less than ϵ, which implies the desired probability is greater than $1 - \epsilon$. This means that a strongly (d, r)-disjunct matrix can be efficiently constructed with probability as high as desired.

Thirdly, we will apply the Lovász Local Lemma to derive an upper bound on the number of rows of a strongly (d, r)-disjunct matrix. The Lovász Local

Lemma first proved by Erdős and Lovász is a powerful tool to prove the existence of combinatorial structures satisfying a prescribed collection of criteria. Here is the lemma in a symmetric form.

Lemma 3. *Let $A_1, A_2, \cdots, A_N$ be events in an arbitrary probability space with $P(A_i) \leq p$ for all $1 \leq i \leq N$. Suppose that each event is mutually independent of all the other events except for at most μ of them. If $ep(\mu+1) \leq 1$, then*

$$P\left(\bigcap_{i=1}^{N} \overline{A_i}\right) > 0,$$

where e denotes the base of natural logarithms.

Let $M = (m_{ij})$ be a random $t_3 \times n$ $0-1$ matrix with $P(m_{ij} = 1) = \frac{1}{d+r}$, $P(m_{ij} = 0) = 1 - \frac{1}{d+r}$, and the entries m_{ij} are mutually pairwise independent. For any two disjoint sets $\mathcal{C}_1$ and $\mathcal{C}_2$ of columns with $|\mathcal{C}_1| = d$ and $|\mathcal{C}_2| = r$, let $A_{\mathcal{C}_1,\mathcal{C}_2}$ be the event that $(\mathcal{C}_1,\mathcal{C}_2)$ is unsatisfied. Obviously,

$$\dot{\mathrm{P}}(A_{\mathcal{C}_1,\mathcal{C}_2}) = \left[1 - \frac{r}{d+r} \cdot (\frac{d+r-1}{d+r})^{d+r-1}\right]^{t_3}$$

and

$$\mu + 1 = \left[\binom{n}{d+r}\binom{d+r}{d} - \binom{n-d-r}{d+r}\binom{d+r}{d}\right].$$

Setting

$$t_3 = \frac{\ln\left[\binom{n}{d+r}\binom{d+r}{d} - \binom{n-d-r}{d+r}\binom{d+r}{d}\right] + 1}{-\ln(1 - \frac{r}{d+r} \cdot (\frac{d+r-1}{d+r})^{d+r-1})}, \tag{7}$$

we have $ep(\mu+1) \leq 1$, and thus by Lemma 3 the desired probability is greater than 0. Consequently, there exists a strongly (d, r)-disjunct matrix of size $t_3 \times n$.

Remark: the purpose of the subsection is to exploit three known methods to derive upper bounds on the length of selectors and to compare the obtained bounds with the one by De Bonis et al. which is the best known result in general cases. Although the obtained results are not as strong as we like (they are asymptotically same), from another point of view, following a conventional analysis in the survey [23], the rate $\lim_{t\to\infty} \frac{\log n}{t}$ derived from the t_1 bound can be shown the best one among all the other ones.

The comparison in the following is based on their original forms, not simplified forms (asymptotic forms). The motivation is in calling awareness to the existence of better results and the alternative constructions (randomness approach).

We now present numerical results of the bounds proposed in this section and that of De Bonis et al. [6] with some parameters. For the sake of fairness, the bound of De Bonis et al. we use for comparison is the original form

$$t_s = \frac{\binom{n}{n/k}}{(k-m+1)\binom{n-k}{n/k-1}} \left[\ln\left(\binom{k-1}{k-m}\binom{n/k}{1}\binom{n-n/k}{k-1}\right) + 1\right],$$

where $k = d+r$ and $m = d+1$ in our setting of strongly (d, r)-disjunct matrices, in the proof in [6, Theorem 1]. The following tables present some numerical results to compare the bounds for some specific parameters. The numerical results show that our constructions are slightly better than the currently best known result by De Bonis et al. [6], at least for the considered parameters.

Table 1 lists the number t of rows obtained by the proposed methods for the case of $n = 300$ and $d = 3$. As shown, setting r an integer close to d seems to be the best choice to get a small number of required tests.

Table 1. The number of rows needed for fixed parameters $n = 300$ and $d = 3$

n	d	r	t_1	t_2	t_3	t_s
300	3	1	175	188	171	186
300	3	2	142	145	137	153
300	3	3	138	136	131	148
300	3	4	140	136	132	150
300	3	5	145	138	136	155
300	3	6	152	142	141	161

Table 2 lists the number of rows required for some small parameters by setting $r = d$. Given a fixed n, the t_2 bound is the best bound for large d, while the t_3 bound is the best for small d except for the case $d = 1$. Notice that the bounds t_2 and t_3 always yield better results than the t_s bound by De Bonis et al.

Table 2. The number of rows required for the $r = d$ case

n	d	r	t_1	t_2	t_3	t_s	min.
300	1	1	27	40	28	44	t_1
300	2	2	84	90	82	98	t_3
300	6	6	283	258	259	278	t_2
300	7	7	327	295	297	316	t_2
1000	1	1	31	48	32	54	t_1
1000	2	2	99	111	98	122	t_3
1000	5	5	287	276	270	302	t_3
1000	12	12	666	603	604	657	t_2
1000	13	13	716	646	647	702	t_2
1000	20	20	1047	926	930	1001	t_2

Acknowledgement. The authors would like to thank the anonymous reviewers for their careful reading and valuable comments.

References

1. Alon, N., Asodi, V.: Learning a hidden subgraph. SIAM J. Discrete Math. 18, 697–712 (2005)
2. Alon, N., Beigel, R., Kasif, S., Rudich, S., Sudakov, B.: Learning a hidden matching. SIAM J. Comput. 33, 487–501 (2004)
3. Barillot, E., Lacroix, B., Cohen, D.: Theoretical analysis of library screening using an n-dimensional pooling strategy. Nucleic Acids Research, 6241–6247 (1991)
4. Berger, T., Mehravari, N., Towsley, D., Wolf, J.: Random multipleaccess communication and group testing. IEEE Trans. Commun. 32, 769–779 (1984)
5. De Bonis, A.: New combinatorial structures with applications to efficient group testing with inhibitors. J. Combin. Optim. 15, 77–94 (2008)
6. De Bonis, A., Gąsieniec, L., Vaccaro, U.: Optimal two-stage algorithms for group testing problems. SIAM J. Comput. 34, 1253–1270 (2005)
7. De Bonis, A., Vaccaro, U.: Optimal algorithms for two group testing problems, and new bounds on generalized superimposed codes. IEEE Trans. Inform. Theory 52, 4673–4680 (2006)
8. Chang, H.L., Chen, H.B., Fu, H.L.: Identification and classification problems on pooling designs for inhibitor models. J. Comput. Biol. 17(7), 927–941 (2010)
9. Chang, H.L., Chen, H.B., Fu, H.L., Shi, C.H.: Reconstruction of hidden graphs and threshold group testing. J. Combin. Optim. 22, 270–281 (2011)
10. Chen, H.B., Du, D.Z., Hwang, F.K.: An unexpected meeting of four seemingly unrelated problems: graph testing, DNA complex screening, superimposed codes and secure key distribution. J. Combin. Optim. 14, 121–129 (2007)
11. Chen, H.B., Fu, H.L.: Nonadaptive algorithms for threshold group testing. Discrete Appl. Math. 157(7), 1581–1585 (2009)
12. Chen, H.B., Fu, H.L., Hwang, F.K.: An upper bound of the number of tests in pooling designs for the error-tolerant complex model. Optim. Lett. 2, 425–431 (2008)
13. Cheng, Y.X., Du, D.Z.: Efficient constructions of disjunct matrices with applications to DNA library screening. J. Comput. Biol. 14, 1208–1216 (2007)
14. Cheng, Y.X., Du, D.Z.: New constructions of one- and two-stage pooling designs. J. Comput. Biol. 15(2), 195–205 (2008)
15. Cheraghchi, M.: Improved Constructions for Non-adaptive Threshold Group Testing. In: Abramsky, S., Gavoille, C., Kirchner, C., Meyer auf der Heide, F., Spirakis, P.G. (eds.) ICALP 2010. LNCS, vol. 6198, pp. 552–564. Springer, Heidelberg (2010)
16. Damaschke, P.: Randomized group testing for mutually obscuring defectives. Inf. Process. Lett. 67, 131–135 (1998)
17. Damaschke, P.: Threshold Group Testing. In: Ahlswede, R., Bäumer, L., Cai, N., Aydinian, H., Blinovsky, V., Deppe, C., Mashurian, H. (eds.) Information Transfer and Combinatorics. LNCS, vol. 4123, pp. 707–718. Springer, Heidelberg (2006)
18. Du, D.Z., Hwang, F.K.: Combinatorial Group Testing and Its Applications, 2nd edn. World Scientific (2000)
19. Du, D.Z., Hwang, F.K.: Pooling Designs and Nonadaptive Group Testing - Important Tools for DNA Sequencing. World Scientific (2006)
20. D'yachkov, A.G.: Superimposed designs and codes for nonadaptive search of mutually obscuring defectives. In: Proc. 2003 IEEE Int. Symp. Inf. Theory, p. 134 (2003)
21. D'yachkov, A.G., Rykov, V.V.: Bounds on the length of disjunct codes. Problemy Peredachi Inform. 18(3), 7–13 (1982)

22. D'yachkov, A.G., Rykov, V.V.: A survey of superimposed code theory. Problems Control Inform. Theory 12, 229–242 (1983)
23. D'yachkov, A.G.: Lectures on designing screening experiments, Lecture Note Series 10, pages (monograph, pp. 112), Combinatorial and Computational Mathematics Center, Pohang University of Science and Technology (POSTECH), Korea Republic (2004)
24. D'yachkov, A.G., Rykov, V.V.: On superimposed codes. In: Fourth International Workshop Algebraic and Combinatorial Coding Theory, Novgorod, Russia, pp. 83–85 (1994)
25. D'yachkov, A.G., Rykov, V.V., Antonov, M.G.: New bounds on rate of the superimposed codes. In: The 10th All-Union Symposium for the Redundancy Problem in Information Systems, Papers, Part 1, St-Petersburg (1989)
26. Farach, M., Kannan, S., Knill, E., Muthukrishnan, S.: Group testing problem with sequences in experimental molecular biology. In: Proc. Compression and Complexity of Sequences, pp. 357–367 (1997)
27. Hong, E.H., Ladner, R.E.: Group testing for image compression. IEEE Trans. Image Process. 11, 901–911 (2002)
28. Hong, Y.W., Scaglione, A.: On multiple access for distributed dependent sensors: a content-based group testing approach. In: IEEE Information Theory Workshop, pp. 298–303 (2004)
29. Hwang, F.K., Sós, V.T.: Nonadaptive hypergeometric group testing. Studia Scient. Math. Hungarica 22 (1987)
30. Kautz, W.H., Singleton, R.R.: Nonrandom binary superimposed codes. IEEE Trans. Inform. Theory 10, 363–377 (1964)
31. Pevzner, P.A., Lipshutz, R.: Towards DNA Sequencing Chips. In: Privara, I., Ružička, P., Rovan, B. (eds.) MFCS 1994. LNCS, vol. 841, pp. 143–158. Springer, Heidelberg (1994)
32. Porat, E., Rothschild, A.: Explicit non-adaptive combinatorial group testing schemes. In: Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP), pp. 748–759 (2008)
33. Stinson, D.R., Wei, R.: Generalized cover-free families. Discrete Math., 463–477 (2004)
34. Stinson, D.R., Wei, R., Zhu, L.: Some new bounds for cover-free families. J. Combin. Theory Ser. A, 224–234 (2000)
35. Torney, D.C.: Sets pooling designs. Ann. Combin., 95–101 (1999)
36. Vermeirssen, V., Deplancke, B., Barrasa, M.I., Reece-Hoyes, J.S., et al.: Matrix and steiner-triple-system smart pooling assays for high-performance transcription regulatory network mapping. Nature Methods 4, 659–664 (2007)
37. Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., et al.: High-quality binary protein interaction map of the yeast interactome network. Science 322, 104–110 (2008)