國立交通大學

應用數學系

碩 士 論 文

羅瓦脣局部引理在匯集設計上的應用

Applications of the Lovász Local Lemma to Pooling Designs

研 究 生：余國安

指導教授：傅恆霖 教授

中 華 民 國 九 十 六 年 六 月

羅瓦茲局部引理在匯集設計上的應用

Applications of the Lovász Local Lemma to Pooling Designs

研 究 生：余國安　　Student：Kuo-An Yu

指導教授：傅恆霖　　Advisor：Hung-Lin Fu

國 立 交 通 大 學

應 用 數 學 系

碩 士 論 文

A Thesis
Submitted to Department of Applied Mathematics
College of Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Master
In

Applied Mathematics

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

謝誌

首先感謝我的指導老師傅恆霖教授，他扮演的是一個慈父的角色，在賦予我獨立研究權力的同時，總是適時地提供我需要的資源，今天的我之所以可以長出許多能力，有一大半是因為他對我的信任以及他給我的安全感。

接著我想感謝陳秋媛教授，相對於傅老師的慈父，秋媛姐扮演慈母的角色，在我就讀交大碩士班的時期，提供許多情感上的支援。

翁志文教授是我見過上課最為完備的一位老師，授課教材總是從最基礎的東西開始，一步一步往前走，會用到的重要工具他一定給出完整的證明，從他那邊我學到做研究如何大膽地假設，小心地求證。

黃大原教授引領我踏入機率方法的領域，在我的論文裡用到的主要工具：羅瓦胥局部引理，就是在他的課堂上學到的。

接著我要感謝的是陳宏賓學長，他宛如我的第二位指導老師，在我徬徨不知做研究為何物的時候，指引我方向，釐清我一些不切實際的想法，讓我的研究可以一點一滴地累積，另外，在我修改論文的時期，毫不吝惜指出我需要改進的地方，如果今天有人稱讚我的論文寫得好，他的指導會是一個很重要的因素。

然後我想要感謝羅元勳學長，協助我突破在(k,m,n)-選擇器的研究上遇到的瓶頸；張惠蘭學姊給了我許多關於口試投影片的建議，讓我的口試進行地更加順利；我的房東王柏森先生提供我住宿的資源，讓我在做研究的路上沒有後顧之憂；好友詹宗智先生提供我電腦硬體與軟體上的協助，讓我可以在編輯論文的工作上不受到環境的限制。

還有好多好多的人要感謝，包括我的同學張雞肉、RE 美女、帥哥陳博士、胖子吳、發誓先生、聯誼王子圳圳、妙妙學姊、育慈、筱凡、兆涵、小強、威雄、排球神童丘小傑…等等，感謝他們陪伴我度過美好的研究生時期，還有人本教育基金會的友人們，一直充實著我的精神生活…，族繁不及備載，謹在此一併致謝。

最後，我要感謝我的爸爸媽媽，感謝他們扶養我長成這樣一個美好的人。

羅瓦胥局部引理在匯集設計上的應用

研究生：余國安　　指導教授：傅恆霖

國 立 交 通 大 學

應 用 數 學 系

摘　要

在計算分子生物學的應用中，一個群試演算法(group testing algorithm)被稱作一個匯集設計(pooling design)，而每個合成測試被稱作一個匯集(pool)。匯集的數目反映了我們必須花費在實驗上的時間與金錢；因此，在測試物件數目固定的前提之下，不管使用逐步演算法(sequential algorithm)或是非調整型演算法(nonadaptive algorithm)，讓匯集的數目最小化是研究群試演算法的最重要任務。

在這篇論文裡，我們主要針對幾類可以應用在匯集設計的矩陣（其中包括(d, r]-分離矩陣、(d, r)-分離矩陣、(d, s out of r]-分離矩陣以及(k, m, n)-選擇器），在固定行(column)數的前提之下，利用羅瓦胥局部引理分別去求得這些矩陣的最小列(row)數的上界。

Applications of the Lovász Local Lemma to Pooling Designs

Student: Kuo-An Yu
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, Taiwan 30050

Advisor: Hung-Lin Fu
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, Taiwan 30050

Abstract

In the application to computational molecular biology, a group testing algorithm is called a *pooling design* and the composition of each test is called a *pool*. The number of tests (pools) reflects to the time and cost we have to spend on the experiment. Therefore, minimizing the number of tests with fixed number of items in either sequential or nonadaptive algorithms is the most important task in the study of group testing algorithms.

In this thesis, we mainly apply the Lovász Local Lemma to obtain upper bounds for the minimum number of rows for $(d,r]$-disjunct matrices, (d,r)-disjunct matrices, $(d, s$ out of $r]$-disjunct matrices, and (k,m,n)-selectors with n columns, respectively, i.e., upper bounds for $t(n,d,r]$, $t(n,d,r)$, $t(n,d,r,s]$, and $t_s(k,m,n)$, respectively, which are listed in the following:

$$t(n,d,r] \leq \left(1+\frac{d}{r}\right)^r \cdot \left(1+\frac{r}{d}\right)^d \cdot \left\{1+\ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]\right\},$$

$$t(n,d,r) \leq \left(1+\frac{d}{r}\right) \cdot \left(1+\frac{r}{d}\right)^{\frac{d}{r}} \cdot \left\{1+\ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]\right\},$$

$$t(n,d,r,s] \leq \frac{1+\ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{f_{d,r,s}(p)}$$

for all $0 < p < 1$, where

$$f_{d,r,s}(p) = (1-p)^d \cdot \left[1 - \sum_{i=0}^{s-1}\binom{r}{i}p^i(1-p)^{r-i}\right],$$

$$t_s(k,m,n) \leq \frac{m}{\binom{k}{m}\cdot m!} \cdot \left[k \cdot \left(1+\frac{1}{k-1}\right)^{k-1}\right]^m \cdot \left\{1+\ln\left[\binom{n}{k} - \binom{n-k}{k}\right]\right\}.$$

Contents

Abstract (in Chinese) i

Abstract (in English) iii

Contents iv

1 Introduction **1**

2 Preliminaries **4**

2.1 Nonadaptive Group Testing 4

2.2 Models 5

2.2.1 The Inhibitor Model 5

2.2.2 The Complex Model 6

2.2.3 The Threshold Model 6

2.3 The Probabilistic Method 6

2.4 The Lovász Local Lemma 7

3 Known Results **9**

3.1 d-Disjunct Matrices 9

3.2 (k, m, n)-Selectors 11

4 Main Results **15**

4.1 $(d, r]$-Disjunct Matrices 15

4.2 (d, r)-Disjunct Matrices 18

4.3 $(d, s$ out of $r]$-Disjunct Matrices 20

4.4 (k, m, n)-Selectors 22

5 Conclusion **24**

Chapter 1

Introduction

Combinatorial group testing is a basic tool in conducting experiments of tests which can be applied to computational molecular biology. A brief description of the basic model is as follows: Given a set N of n items consisting of at most d positive (used to be called *defective*) items with the others being negative (used to be called *good*). Let P denote the set of all positive items. The problem is to identify P. The tool of identification is the so-called group tests, sometimes called *pools*, while a group test is applicable to an arbitrary subset S of N with two possible outcomes; a negative outcome indicates that all items in S are negative; a positive outcome indicates that there are at least one positive item in S, not knowing which one or how many. The goal is to minimize the number of such tests in identifying P.

Li [12] started to consider combinatorial group testing where the presumed knowledge on the set of defectives is that it must be a member, called a *sample*, of a given family called a *sample space*. For instance, the sample space could consist of all d-subsets of the n items when the presumed knowledge is that there are exactly d defectives among the n items. We will refer to this space as the $S(d, n)$ space while the $S(\overline{d}, n)$ space specifies that d is an upper bound of the number of defectives.

Group testing algorithms (*pooling designs*) can be generally divided into two types: *sequential* and *nonadaptive*. A sequential algorithm conducts the tests one by one and the outcomes of all previous tests can be used to set up the later test. A nonadaptive algorithm specifies all tests in advance so that they can be conducted simultaneously; thus forbidding using the information of previous tests to design later ones. In most applications to molecular biology, an experiment can be time-consuming. Therefore, it is much preferable to have a nonadaptive algorithm. In this thesis we′ll focus on some matrices used for nonadaptive group testing.

There are various models for group testing. In screening clone library, the goal is to determine which clones in the library hybridize with a given probe in an efficient fashion. A clone is said to be positive if it hybridizes with the given probe, and negative otherwise. In practical applications, there is another category of clones besides positive and negative clones, called *inhibitors* whose effect is to neutralize positive clones. Therefore, we shall have models of group testing with or without inhibitors. Also in applications, we may face the situation that the property of being positive or negative is defined on subsets of items instead of on individual items. Such models are known as *complex models*. The study of complex models does have a significant impact in recent years. As a generalization of the classical group testing problem, the *threshold model* appears.

The probabilistic method is a useful tool for tackling many problems in discrete mathematics. Roughly speaking, the method works as follows. Trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability. Among various probabilistic methods, the Lovász Local Lemma, first proved by Erdős and Lovász [10], is extremely powerful and plays the main role in this thesis.

In this thesis, we first introduce a few types of matrices such as separable or disjunct matrices and also the relationship between them and nonadaptive group testing. Then we introduce some models for group testing. Next, also in Chapter 2, we illustrate the probabilistic method by a simple example, followed by reviewing our main tool, the Lovász Local Lemma. In Chapter 3, we review two known results: d-disjunct matrices by Yeh [17] and (k, m, n)-selectors by De Bonis, Gąsieniec, and Vaccaro [5]. Finally, in Chapter 4, we obtain various upper bounds for the minimum number of rows for $(d, r]$-disjunct matrices, (d, r)-disjunct matrices, $(d, s$ out of $r]$-disjunct matrices, and (k, m, n)-selectors with n columns, respectively, by applying the Lovász Local Lemma.

Chapter 2

Preliminaries

2.1 Nonadaptive Group Testing

A nonadaptive group testing algorithm can be represented by a binary matrix $M = (m_{ij})$ where rows are indexed by pools, columns by items, and $m_{ij} = 1$ if and only if item j is in pool i. For convenience, we identify a column C_j of M with a set of row indices corresponding to the 1-entries in C_j. Hence we could consider union or intersection of some columns of M. In the classic group testing problem, three types of binary matrices have been the major tools in understanding and constructing a pooling design.

Definition 2.1.1. A $t \times n$ binary matrix M is called d-separable if for any two distinct d-sets D, D' of columns of M, $\bigcup D \neq \bigcup D'$, i.e., no two unions of d columns of M are the same.

Definition 2.1.2. A $t \times n$ binary matrix M is called $\overline{d}$-separable if for any two distinct sets D, D' of columns of M with $|D|, |D'| \leq d$, $\bigcup D \neq \bigcup D'$, i.e., no two unions of at most d columns of M are the same.

Definition 2.1.3. A $t \times n$ binary matrix M is called d-disjunct if the union of any d columns does not contain any other column in M.

We explain the properties in the above definitions in terms of pooling designs. Consider the sample space $S(d, n)$ where exact d positive items are present. The d-separability property shows that each sample in $S(d, n)$ induces a different outcome vector. Hence there is a 1-1 correspondence between outcome vectors and samples in $S(d, n)$, and the d positive items can be identified. Moreover, the d-separability is also a necessary condition for a matrix M used to identify the d positive items. Similarly, the $\overline{d}$-separability shows that samples in $S(\overline{d}, n)$, where at most d positive items are present, are distinguishable while the d-disjunctness guarantees an appearing of each negative item in some negative pool.

2.2 Models

More detailed descriptions of some models for group testing are given in this section.

2.2.1 The Inhibitor Model

In some applications, an item can be positive, negative, or anti-positive in the sense that the presence of anti-positives cancels the effect of positives. They are called inhibitors in the literature. In the simplest inhibitor model, first proposed by Farach et al. [11], the presence of an inhibitor in a pool dictates a negative outcome, regardless of the presence of positive items in the pool.

Consider a set N of n items consisting of at most d positives and at most h inhibitors with the others being negatives. Let P denote the set of all positive items and I the set of all inhibitors. The usual concern in the inhibitor model is to identify the set P. Another interesting problem one can consider is to also identify the inhibitor set I.

2.2.2 The Complex Model

In the complex model, we consider a set N of n items and an unknown family $P = \{P_i\}$ of subsets of N where each such subset is a cause of a certain given biological phenomenon. A set S of items which is a candidate of a member of P is called a *complex* while members of P are called positive complexes. The problem is to identify P from a given set of complexes. An experiment can be applied to an arbitrary complex S with two possible outcomes; a positive outcome indicates that S contain some $P_i \in P$, while a negative outcome indicates the remaining cases.

2.2.3 The Threshold Model

The threshold model is quite a natural generalization of the classical group testing problem, which is described as follows. Consider a set N of n items containing a set P of positive items with the others being negative. Let l and u be two nonnegative integers with $l < u$, called the lower and upper threshold, respectively. A group test applied to a subset S of items shows positive if S contains at least u positives, and negative if at most l positives are present in S. If the number of positives in S is between l and u, the test will show an arbitrary answer. The goal is still to identify P. Clearly, the classic group testing problem is a special case of the threshold model with $l = 0$ and $u = 1$.

2.3 The Probabilistic Method

We illustrate the probabilistic method by a simple example, which is presented in Alon and Spencer [1].

The Ramsey number $R(k, l)$ is the smallest integer n such that in any two-coloring of the edges of a complete graph K_n on n vertices by red and blue, either there is a red K_k (i.e., a complete subgraph on k vertices all of whose edges are colored red) or there is a blue K_l. Ramsey [14] showed that $R(k,l)$ is finite for any two integers k and l. Here, we show that if $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$, then $R(k,k) > n$. Thus $R(k,k) > \left\lfloor 2^{\frac{k}{2}} \right\rfloor$ for all $k \geq 3$. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For any fixed set R of k vertices, let A_R be the event that the induced subgraph of K_n on R is monochromatic (i.e., that either all its edges are red or they are all blue). Clearly, $Pr(A_R) = 2^{1-\binom{k}{2}}$. Since there are $\binom{n}{k}$ possible choices for R, the probability that at least one of the events A_R occurs is at most $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$. Thus, with positive probability, no event A_R occurs and there is a two-coloring of K_n without a monochromatic K_k, i.e., $R(k,k) > n$. Note that if $k \geq 3$ and we take $n = \left\lfloor 2^{\frac{k}{2}} \right\rfloor$, then

$$\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{\frac{k^2}{2}}} < 1$$

and hence $R(k,k) > \left\lfloor 2^{\frac{k}{2}} \right\rfloor$ for all $k \geq 3$.

2.4 The Lovász Local Lemma

There is a trivial case in which one can show that a certain event holds with positive, though small, probability. Indeed, if we have n mutually independent events and each of them holds with probability at least $p > 0$, then the probability that all events hold simultaneously is at least p^n, which is positive, although it may be exponentially small in n. It is natural to expect that the case of mutual independence can be generalized to that of rare dependencies, and provide a more general way of proving that certain events hold with positive, though small, probability. Such a generalization is indeed possible and is stated in the Lovász Local Lemma.

Next, we review the main ideas of the Lovász Local Lemma, following the treatment described in Alon and Spencer [1].

Definition 2.4.1. Let $A_1, A_2, \cdots, A_n$ be events in an arbitrary probability space. A graph $G = (V, E)$ on the set of vertices $V = \{1, 2, \cdots, n\}$ is said to be a dependency graph for the events $A_1, A_2, \cdots, A_n$ if for each $i, 1 \le i \le n$, the event A_i is mutually independent of a set of all the other events except for those A_j with $\{i, j\} \in E$.

We're now in the position to state the Lovász Local Lemma by skipping its proof:

Theorem 2.4.2. *(The Lovász Local Lemma; General Case)*
Let $A_1, A_2, \cdots, A_n$ be events in an arbitrary probability space and let $G = (V, E)$ be a dependency graph for them. Suppose there are real numbers $x_1, \cdots, x_n$ such that $0 \le x_i < 1$ and $Pr(A_i) \le x_i \cdot \prod_{\{i,j\}\in E}(1 - x_j)$ for all $1 \le i \le n$. Then $Pr(\bigcap_{i=1}^{n} \overline{A_i}) \ge \prod_{i=1}^{n}(1 - x_i)$. In particular, with positive probability no event A_i holds.

The next corollary establishes a result that holds when all events have probability at most p, for some constant p. In this corollary and elsewhere, e denotes the base of natural logarithms (i.e., $e \approx 2.71828$).

Corollary 2.4.3. *(The Lovász Local Lemma; Symmetric Case)*
Let $A_1, A_2, \cdots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most μ, and that $Pr(A_i) \le p$ for all $1 \le i \le n$. If $e \cdot p \cdot (\mu + 1) \le 1$, then $Pr(\bigcap_{i=1}^{n} \overline{A_i}) > 0$.

In the remaining of this thesis, our goal is to prove the existence of some kind of matrix with the desired properties under some conditions, e.g., the number of rows is large enough, by using Corollary 2.4.3. Thus deducing an upper bound for the minimum size of this kind of matrix.

Chapter 3

Known Results

3.1 d-Disjunct Matrices

For positive integers n and d, let $[n]$ denote the set $\{1, 2, \cdots, n\}$ and $\binom{[n]}{d}$ denote the collection of all subsets of $[n]$ with cardinality d. Let $t(d, n)$ denote the minimum number of rows for a d-disjunct matrix with n columns. Yeh [17] proves the following theorem by using Corollary 2.4.3. For completeness, we include his proof in what follows, with a little adjustment.

Theorem 3.1.1. *[17]*

$$t(d,n) \leq (d+1) \cdot \left(1 + \frac{1}{d}\right)^d \cdot \left\{1 + \ln\left[(d+1)\left(\binom{n}{d+1} - \binom{n-d-1}{d+1}\right)\right]\right\}.$$

Proof. Let $M = (m_{ij})$ be a $\frac{t}{q} \times n$ random matrix with entries in $\{1, 2, \cdots, q\}$ such that $Pr(m_{ij} = k) = \frac{1}{q}$ for $1 \leq k \leq q$, and the entries m_{ij} are mutually independent. Let M^* be a $t \times n$ random $\{0, 1\}$-matrix converted from M by replacing each q-ary alphabet by a unique q-digit binary column array with unit weight. For example, when $q = 3$, the replacement can be

$$1 \rightarrow \begin{matrix} 1 \\ 0 \\ 0 \end{matrix}, \quad 2 \rightarrow \begin{matrix} 0 \\ 1 \\ 0 \end{matrix}, \quad 3 \rightarrow \begin{matrix} 0 \\ 0 \\ 1 \end{matrix}.$$

Let $C_1, \cdots, C_n$ be the columns of M^*. For $J \in \binom{[n]}{d}$ and $s \in [n] \backslash J$, let $A_{J,s}$ be the event that the union of columns C_j, $j \in J$, contains column C_s. For $i \in \left[\frac{t}{q}\right]$, let $A_{i,J,s}$ be the event that $m_{ij} = m_{is}$ for some $j \in J$. Then

$$\begin{aligned} Pr(A_{J,s}) &= Pr\left(\bigcap_{i=1}^{\frac{t}{q}} A_{i,J,s}\right) = \prod_{i=1}^{\frac{t}{q}} Pr(A_{i,J,s}) \\ &= \left[1 - Pr(\overline{A_{i,J,s}})\right]^{\frac{t}{q}} \\ &= \left[1 - q \cdot \frac{1}{q} \cdot \left(1 - \frac{1}{q}\right)^d\right]^{\frac{t}{q}} \\ &= \left[1 - \left(1 - \frac{1}{q}\right)^d\right]^{\frac{t}{q}}. \end{aligned}$$

Note that $A_{J,s}$ is mutually independent of all the other events $A_{J',s'}$ except for those with $(J' \cup \{s'\}) \cap (J \cup \{s\}) \neq \phi$. There are exactly

$$(d+1) \cdot \left[\binom{n}{d+1} - \binom{n-d-1}{d+1}\right] - 1$$

such events. According to Corollary 2.4.3, a $t \times n$ d-disjunct matrix exists whenever

$$e \cdot \left[1 - \left(1 - \frac{1}{q}\right)^d\right]^{\frac{t}{q}} \cdot (d+1) \cdot \left[\binom{n}{d+1} - \binom{n-d-1}{d+1}\right] \leq 1.$$

holds. Taking natural logarithm to both sides yields the equivalent inequality

$$t \geq q \cdot \frac{1 + \ln\left[(d+1) \cdot \left(\binom{n}{d+1} - \binom{n-d-1}{d+1}\right)\right]}{-\ln\left[1 - \left(1 - \frac{1}{q}\right)^d\right]}. \tag{3.1}$$

Using the fact that $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that whenever the inequality

$$t \geq q \cdot \frac{1 + \ln\left[(d+1) \cdot \left(\binom{n}{d+1} - \binom{n-d-1}{d+1}\right)\right]}{\left(1 - \frac{1}{q}\right)^d} \tag{3.2}$$

holds, (3.1) holds. To minimize the R.H.S. of (3.2), we let $q = d+1$ and complete the proof. ∎

3.2 (k, m, n)-Selectors

We begin this section with the definition of a (k, m, n)-selector.

Definition 3.2.1. Given integers k, m, and n, with $1 \leq m \leq k \leq n$, we say that a $t \times n$ binary matrix M is a (k, m, n)-selector if any submatrix of M obtained by choosing k out of n arbitrary columns of M contains at least m distinct rows of the identity matrix I_k. The integer t is the size of the (k, m, n)-selector.

As the relationship between (k, m, n)-selectors and group testing, De Bonis, Gąsieniec, and Vaccaro [5] proved that there exists a two-stage group testing algorithm for finding up-to-d positives out of n items and that uses a number of tests equal to $t + k - 1$, where t is the size of a $(k, d + 1, n)$-selector.

Let $t_s(k, m, n)$ denote the minimum size of a (k, m, n)-selector. De Bonis, Gąsieniec, and Vaccaro [5] obtain upper bounds for $t_s(k, m, n)$ by translating the problem into the hypergraph language. Still for completeness, we include their proof in what follows. Given a finite set X and a family $\mathcal{F}$ of subsets of X, a hypergraph is a pair $\mathcal{H} = (X, \mathcal{F})$. Elements of X will be called vertices of $\mathcal{H}$, and elements of $\mathcal{F}$ will be called hyperedges of $\mathcal{H}$. A cover of $\mathcal{H}$ is a subset $T \subseteq X$ such that for any hyperedge $E \in \mathcal{F}$ we have $T \cap E \neq \phi$. The minimum size of a cover of $\mathcal{H}$ will be denoted by $\tau(\mathcal{H})$. A fundamental result by Lovász [13] implies that

$$\tau(\mathcal{H}) < \frac{|X|}{\min_{E \in \mathcal{F}} |E|}(1 + \ln \Delta), \tag{3.3}$$

where $\Delta = \max_{x \in X} |\{E : x \in E \in \mathcal{F}\}|$.

Essentially, Lovász proves that, by greedily choosing vertices in X that intersect the maximum number of yet nonintersected hyperedges of $\mathcal{H}$, one obtains a cover of a size smaller than the R.H.S. of (3.3). Our aim is to show that (k, m, n)-selectors are covers of properly defined hypergraphs. Lovász's result (3.3) will then provide us with the desired upper bound on the minimum selector size.

We shall proceed as follows. Let X be the set of all binary vectors $\mathbf{x} = (x_1, \cdots, x_n)$ of length n containing n/k 1's (the value n/k is a consequence of an optimized choice whose justification can be skipped here). For any integer i, $1 \leq i \leq k$, denote by $\mathbf{a}_i$ the binary vector of length k having all components equal to zero with the exception of the component in position i. Moreover, for any set of indices $S = \{i_1, \cdots, i_k\}$, with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and for any binary vector $\mathbf{a} = (a_1, \cdots, a_k) \in \{\mathbf{a}_1, \cdots, \mathbf{a}_k\}$, define the set of binary vectors $E_{\mathbf{a},S} = \{\mathbf{x} = (x_1, \cdots, x_n) \in X : x_{i_1} = a_1, \cdots, x_{i_k} = a_k\}$. For any set $A \subseteq \{\mathbf{a}_1, \cdots, \mathbf{a}_k\}$ of size r, $r = 1, \cdots, k$, and any set $S \subseteq \{1, \cdots, n\}$ with $|S| = k$, define $E_{A,S} = \bigcup_{\mathbf{a} \in A} E_{\mathbf{a},S}$. For any $r = 1, \cdots, k$ we define $\mathcal{F}_r = \{E_{A,S} : A \subset \{\mathbf{a}_1, \cdots, \mathbf{a}_k\}, |A| = r, S \subseteq \{1, \cdots, n\}, |S| = k\}$ and the hypergraph $\mathcal{H}_r = (X, \mathcal{F}_r)$. We claim that any cover T of $\mathcal{H}_{k-m+1}$ is a (k, m, n)-selector; i.e., any submatrix of k arbitrary columns of T contains at least m distinct rows of the identity matrix I_k. The proof is done by contradiction. Assume that there exists a set of indices $S = \{i_1, \cdots, i_k\}$ such that the submatrix of T obtained by considering only the columns of T with indices $i_1, \cdots, i_k$ contains at most $m - 1$ distinct rows of I_k. Let such rows be $\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}$, with $s \leq m-1$; let A be any subset of $\{\mathbf{a}_1, \cdots, \mathbf{a}_k\} \setminus \{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_s}\}$ of cardinality $|A| = k - m + 1$; and let $E_{A,S}$ be the corresponding hyperedge of $\mathcal{H}_{k-m+1}$. By construction we have that $T \cap E_{A,S} = \phi$, contradicting the fact that T is a cover for $\mathcal{H}_{k-m+1}$.

The above proof that (k, m, n)-selectors coincide with the covers of $\mathcal{H}_{k-m+1}$ allows us to use Lovász's result (3.3) to give upper bounds for $t_s(k, m, n)$.

Theorem 3.2.2. *[5]*

$$t_s(k,m,n) < \frac{ek^2}{k-m+1}\ln\frac{n}{k} + \frac{ek(2k-1)}{k-m+1}, \tag{3.4}$$

where $e = 2.71828...$ *is the base of the natural logarithm.*

Proof. We need only to evaluate the quantities $|X|$, $\min\{|E| : E \in \mathcal{F}_{k-m+1}\}$, and Δ for the hypergraph $\mathcal{H}_{k-m+1}$. By definition $|X| = \binom{n}{n/k}$. Moreover, each hyperedge $E_{A,S}$ of $\mathcal{H}_{k-m+1}$ is the union of $k-m+1$ disjoint sets $E_{\mathbf{a},S}$; therefore it has cardinality

$$|E_{A,S}| = (k-m+1)\cdot|E_{\mathbf{a},S}| = (k-m+1)\binom{n-k}{n/k-1}.$$

To compute Δ, observe that each $\mathbf{x} \in X$ belongs to $\binom{n/k}{1}\binom{n-n/k}{k-1}$ distinct sets $E_{\mathbf{a},S}$, and each $E_{\mathbf{a},S}$ belongs to $\binom{k-1}{k-m}$ distinct hyperedges $E_{A,S}$. Therefore, for $\mathcal{H}_{k-m+1}$ we have

$$\Delta = \binom{k-1}{k-m}\binom{n/k}{1}\binom{n-n/k}{k-1}.$$

Hence one has

$$t(k,m,n) < \frac{\binom{n}{n/k}}{(k-m+1)\binom{n-k}{n/k-1}}\left[1+\ln\binom{k-1}{k-m}\binom{n/k}{1}\binom{n-n/k}{k-1}\right]. \tag{3.5}$$

For $k \in \{1,2\}$, it is $\dfrac{\binom{n}{n/k}}{\binom{n-k}{n/k-1}} < 2k$, whereas for $k \geq 3$ it is

$$\begin{aligned}
\frac{\binom{n}{n/k}}{\binom{n-k}{n/k-1}} &= k\frac{n-1}{n-n/k} \cdot \frac{n-2}{n-n/k-1} \times \cdots \times \frac{n-k+1}{n-k-n/k+2} \\
&\leq k\left(\frac{n-k+1}{n-k-n/k+2}\right)^{k-1} \\
&= k\left(\frac{k(n-k+1)}{k(n-k+1)-(n-k)}\right)^{k-1} \\
&= k\left(1+\frac{n-k}{k(n-k+1)-(n-k)}\right)^{k-1} \\
&\leq k\left(1+\frac{1}{k-1}\right)^{k-1} \\
&< ek.
\end{aligned}$$

Moreover, using the well-known inequality $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$, one can conclude

$$\begin{aligned}
\binom{k-1}{k-m}\binom{n/k}{1}\binom{n-n/k}{k-1} &\leq \left(\frac{k-1}{k-m}\right)^{k-m} e^{k-m}\frac{n}{k}\left(\frac{n-n/k}{k-1}\right)^{k-1} e^{k-1} \\
&= e^{2k-m-1}\left(1+\frac{m-1}{k-m}\right)^{k-m}\left(\frac{n}{k}\right)^k \\
&\leq e^{2k-m-1}\left(1+\frac{m}{k-m}\right)^{k-m}\left(\frac{n}{k}\right)^k \\
&\leq e^{2k-m-1}e^m\left(\frac{n}{k}\right)^k.
\end{aligned}$$

The theorem now follows from (3.5) and the above inequalities. ∎

Chapter 4

Main Results

4.1 $(d, r]$-Disjunct Matrices

To generalize Theorem 3.1.1, we start by giving a more general definition.

Definition 4.1.1. A $t \times n$ binary matrix M is called $(d, r]$-disjunct if the union of any d columns does not contain the intersection of any other r columns in M. Clearly, $(d, 1]$-disjunctness is precisely d-disjunctness.

As the relationship between $(d, r]$-disjunct matrices and nonadaptive group testing, Chen, Du and Hwang [2] proved that a $(d, r]$-disjunct matrix can identify the up-to-d positives on the complex model.

Let $t(n, d, r]$ denote the minimum number of rows for a $(d, r]$-disjunct matrix with n columns. We have the following generalization of Theorem 3.1.1, followed by the proof using the same approach used in the proof of Theorem 3.1.1.

Theorem 4.1.2.

$$t(n, d, r] \leq \left(1 + \frac{d}{r}\right)^{r} \cdot \left(1 + \frac{r}{d}\right)^{d} \cdot \left\{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]\right\}. \tag{4.1}$$

Proof. Let M and M^* be as in the proof of Theorem 3.1.1. Again let $C_1, \cdots, C_n$ be the columns of M^*. For $D \in \binom{[n]}{d}$ and $R \in \binom{[n]}{r}$ with $D \cap R = \phi$, let $A_{D,R}$ be the event that the union of columns C_j, $j \in D$, contains the intersection of columns C_k, $k \in R$. For $i \in \left[\frac{t}{q}\right]$, let $\overline{A_{i,D,R}}$ be the event that $m_{ik_1} = m_{ik_2}$ for all $k_1 \neq k_2 \in R$ and $m_{ij} \neq m_{ik_1}$ for all $j \in D$. Then

$$\begin{aligned}
Pr(A_{D,R}) &= Pr\left(\bigcap_{i=1}^{\frac{t}{q}} A_{i,D,R}\right) = \prod_{i=1}^{\frac{t}{q}} Pr(A_{i,D,R}) \\
&= \left[1 - Pr(\overline{A_{i,D,R}})\right]^{\frac{t}{q}} \\
&= \left[1 - q \cdot \left(\frac{1}{q}\right)^{r} \cdot \left(1 - \frac{1}{q}\right)^{d}\right]^{\frac{t}{q}} \\
&= \left[1 - \left(\frac{1}{q}\right)^{r-1} \left(1 - \frac{1}{q}\right)^{d}\right]^{\frac{t}{q}}.
\end{aligned}$$

Note that $A_{D,R}$ is mutually independent of all the other events $A_{D',R'}$ except for those with $(D' \cup R') \cap (D \cup R) \neq \phi$. There are exactly

$$\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r} - 1$$

such events. According to Corollary 2.4.3, a $t \times n$ $(d, r]$-disjunct matrix exists whenever

$$e \cdot \left[1 - \left(\frac{1}{q}\right)^{r-1} \left(1 - \frac{1}{q}\right)^{d}\right]^{\frac{t}{q}} \cdot \left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right] \leq 1$$

holds. Taking natural logarithm to both sides yields the equivalent inequality

$$(4.2) \qquad t \geq q \cdot \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{-\ln\left[1 - \left(\frac{1}{q}\right)^{r-1} \left(1 - \frac{1}{q}\right)^{d}\right]}.$$

Using the fact that $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that whenever

$$(4.3) \qquad t \geq q \cdot \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{\left(\frac{1}{q}\right)^{r-1} \left(1 - \frac{1}{q}\right)^{d}}$$

holds, (4.2) holds. To minimize the R.H.S. of (4.3), we let $q = \frac{d}{r} + 1$ and complete the proof. ∎

In the above proof of Theorem 4.1.2, some small problems may occur. For example, with the restriction that the number of rows of M and q must be positive integers, how about q doesn't divide t or r doesn't divide d? For this sake, we provide another proof of Theorem 4.1.2 by omitting the process converting M into M^* and letting M be a random $\{0,1\}$-matrix directly. (Note that in the remaining sections of this chapter, we adopt the above technique.) However, if q divides t and r divides d, the above proof says more: the column sum of the desired matrix equals a constant $\frac{t}{q}$. The following is our second proof of Theorem 4.1.2.

Proof. Let $M = (m_{ij})$ be a $t \times n$ random $\{0,1\}$-matrix with $Pr(m_{ij} = 1) = p$, $Pr(m_{ij} = 0) = 1 - p$, and the entries m_{ij} are mutually independent. Let $C_1, \cdots, C_n$ be the columns of M. For $D \in \binom{[n]}{d}$ and $R \in \binom{[n]}{r}$ with $D \cap R = \phi$, let $A_{D,R}$ be the event that the union of columns C_j, $j \in D$, contains the intersection of columns C_k, $k \in R$. Then

$$Pr(A_{D,R}) = \left[1 - p^r \cdot (1-p)^d\right]^t.$$

Similar to the first proof, a $t \times n$ $(d, r]$-disjunct matrix exists whenever

$$e \cdot \left[1 - p^r \cdot (1-p)^d\right]^t \cdot \left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right] \leq 1$$

holds, which is equivalent to

(4.4) $$t \geq \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{-\ln\left[1 - p^r \cdot (1-p)^d\right]}.$$

Using the fact that $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that whenever

$$t \geq \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{p^r \cdot (1-p)^d} \tag{4.5}$$

holds, (4.4) holds. To minimize the R.H.S. of (4.5), we let $p = \dfrac{r}{d+r}$ and complete the proof. ∎

Chen, Fu and Hwang [3] also provided an upper bound for $t(n, d, r]$:

$$t(n, d, r] < \left(1 + \frac{d}{r}\right)^r \cdot \left(1 + \frac{r}{d}\right)^d \cdot \left\{1 + (d+r) \cdot \left[1 + \ln\left(\frac{n}{d+r} + 1\right)\right]\right\}. \tag{4.6}$$

Observe that the bound in (4.6) is $O((d+r)\ln n)$ and the bound in (4.1) is $O((d+r-1)\ln n)$, which is a little bit better.

Note that Stinson and Wei [16] provided two asymptotic upper bounds for $t(n, d, r]$ by using two other structures. One bound is $O\left(\binom{d+r}{r}(dr)^{\log^* n}\log n\right)$, where the function $\log^*$ is defined recursively by $\log^*(1) = 1$ and $\log^* n = \log^*(\lceil \log n \rceil) + 1$ if $n > 1$. The bound of the other one is $O\left(\binom{d+r}{r}\log n\right)$. Also note that their bounds are asymptotic and our bound in (4.1) is non-asymptotic.

4.2 (d, r)-Disjunct Matrices

We present another generalization of Theorem 3.1.1 in this section.

Definition 4.2.1. A $t \times n$ binary matrix M is called (d, r)-disjunct if the union of any d columns does not contain the union of any other r columns in M. Clearly, $(d, 1)$-disjunctness is precisely d-disjunctness.

As the relationship between (d, r)-disjunct matrices and nonadaptive group testing, De Bonis and Vaccaro [6] proved that the (h, d)-disjunctness is a necessary condition for identifying P on the (d, h)-inhibitor model.

Let $t(n,d,r)$ denote the minimum number of rows for a (d,r)-disjunct matrix with n columns. We have the following generalization of Theorem 3.1.1.

Theorem 4.2.2.

$$(4.7)\quad t(n,d,r) \le \left(1+\frac{d}{r}\right)\cdot\left(1+\frac{r}{d}\right)^{\frac{d}{r}}\cdot \left\{1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]\right\}.$$

Proof. Let $M=(m_{ij})$ be a $t\times n$ random $\{0,1\}$-matrix with $Pr(m_{ij}=1)=p$, $Pr(m_{ij}=0)=1-p$, and the entries m_{ij} are mutually independent. Let $C_1,\cdots,C_n$ be the columns of M. For $D\in\binom{[n]}{d}$ and $R\in\binom{[n]}{r}$ with $D\cap R=\phi$, let $A_{D,R}$ be the event that the union of columns C_j, $j\in D$, contains the union of columns C_k, $k\in R$. Then

$$Pr(A_{D,R})=\left\{1-(1-p)^d\cdot[1-(1-p)^r]\right\}^t.$$

Similar to the proof of Theorem 4.1.2, a $t\times n$ (d,r)-disjunct matrix exists whenever

$$e\cdot\left\{1-(1-p)^d\cdot[1-(1-p)^r]\right\}^t\cdot \left[\binom{n}{d}\binom{n-d}{r}-\binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]\le 1$$

holds, which is equivalent to

$$(4.8)\qquad t\ge\frac{1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{-\ln\left\{1-(1-p)^d\cdot[1-(1-p)^r]\right\}}.$$

Using the fact that $-\ln(1-x)\ge x$ for $0\le x<1$, we conclude that whenever

$$(4.9)\qquad t\ge\frac{1+\ln\left[\binom{n}{d}\binom{n-d}{r}-\binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{(1-p)^d\cdot[1-(1-p)^r]}$$

holds, (4.8) holds. To minimize the R.H.S. of (4.9), we let $p=1-\left(\frac{d}{d+r}\right)^{\frac{1}{r}}$ and complete the proof. ∎

Du and Hwang [9] proved that a (k, m, n)-selector is $(m-1, k-m+1)$-disjunct, which implies that a $(d+r, d+1, n)$-selector is (d, r)-disjunct. By Theorem 3.2.2, we have

$$t(n, d, r) < \frac{e(d+r)^2}{r} \ln \frac{n}{d+r} + \frac{e(d+r)[2(d+r)-1]}{r}. \tag{4.10}$$

Note that the bound in (4.10) is $O((d+r)\ln n)$ and the bound in (4.7) is $O((d+r-1)\ln n)$, which is a little bit better.

4.3 $(d, s$ out of $r]$-Disjunct Matrices

In section 4.1 and 4.2, two versions of generalizations of Theorem 3.1.1 are given. However, there exists a more generalized category containing these two versions, which is presented in this section.

Definition 4.3.1. For $1 \le s \le r$, a $t \times n$ binary matrix M is called $(d, s$ out of $r]$-disjunct if for any d columns and any other r columns of M, there exists a row index in which none of the d columns appear and at least s of the r columns do. Clearly, $(d, 1$ out of $r]$-disjunctness is precisely (d, r)-disjunctness and $(d, r$ out of $r]$-disjunctness is precisely $(d, r]$-disjunctness.

Let $t(n, d, r, s]$ denote the minimum number of rows for a $(d, s$ out of $r]$-disjunct matrix with n columns. We have the following theorem:

Theorem 4.3.2.

$$t(n, d, r, s] \le \frac{1 + \ln \left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r} \right]}{f_{d,r,s}(p)} \tag{4.11}$$

for all $0 < p < 1$, *where*

$$f_{d,r,s}(p) = (1-p)^d \cdot \left[1 - \sum_{i=0}^{s-1} \binom{r}{i} p^i (1-p)^{r-i} \right].$$

Proof. Let $M = (m_{ij})$ be a $t \times n$ random $\{0,1\}$-matrix with $Pr(m_{ij} = 1) = p$, $Pr(m_{ij} = 0) = 1 - p$, and the entries m_{ij} are mutually independent. Let $C_1, \cdots, C_n$ be the columns of M. For $D \in \binom{[n]}{d}$ and $R \in \binom{[n]}{r}$ with $D \cap R = \phi$, let $\overline{A_{D,R}}$ be the event that there exists a row index in which none the columns C_j, $j \in D$, appear and at least s of the columns C_k, $k \in R$, do. Then

$$Pr(A_{D,R}) = \left\{1 - (1-p)^d \cdot \left[1 - \sum_{i=0}^{s-1} \binom{r}{i} p^i (1-p)^{r-i}\right]\right\}^t . \tag{4.12}$$

Define the function

$$f_{d,r,s}(p) = (1-p)^d \cdot \left[1 - \sum_{i=0}^{s-1} \binom{r}{i} p^i (1-p)^{r-i}\right]$$

for $0 < p < 1$. Then (4.12) becomes

$$Pr(A_{D,R}) = [1 - f_{d,r,s}(p)]^t .$$

Similar to the proof of Theorem 4.1.2, a $t \times n$ $(d, s$ out of $r]$-disjunct matrix exists whenever

$$e \cdot [1 - f_{d,r,s}(p)]^t \cdot \left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right] \leq 1$$

holds, which is equivalent to

$$t \geq \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{-\ln\left[1 - f_{d,r,s}(p)\right]}. \tag{4.13}$$

Using the fact that $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that whenever

$$t \geq \frac{1 + \ln\left[\binom{n}{d}\binom{n-d}{r} - \binom{n-(d+r)}{d}\binom{n-(d+r)-d}{r}\right]}{f_{d,r,s}(p)}$$

holds, (4.13) holds, completing the proof. ∎

4.4 (k, m, n)-Selectors

We can use similar approaches to obtain an upper bound for the minimum size of a (k, m, n)-selector.

Theorem 4.4.1.
(4.14)

$$t_s(k,m,n) \le \frac{m}{\binom{k}{m} \cdot m!} \cdot \left[k \cdot \left(1 + \frac{1}{k-1}\right)^{k-1} \right]^m \cdot \left\{ 1 + \ln\left[\binom{n}{k} - \binom{n-k}{k} \right] \right\}.$$

Proof. Let $M^* = (m_{ij})$ be a $t \times n$ random binary matrix with $Pr(m_{ij} = 1) = p$, $Pr(m_{ij} = 0) = 1 - p$, and the entries m_{ij} are mutually independent. For $K \in \binom{[n]}{k}$ and $M \in \binom{[t]}{m}$, define A_K be the event that the $t \times k$ submatrix of M^* corresponding to K contains at most $m - 1$ rows of I_k, and $A_{K,M}$ be the event that the $m \times k$ submatrix of M^* corresponding to K and M doesn't consist of m distinct rows of I_k. Observe that

$$A_K = \bigcap_{M \in \binom{[t]}{m}} A_{K,M}.$$

Let $M_i = \{m \cdot (i-1) + 1, m \cdot (i-1) + 2, \cdots, mi\}$ for $1 \le i \le \frac{t}{m}$. Then

$$\begin{aligned} Pr(A_K) &= Pr\left(\bigcap_{M \in \binom{[t]}{m}} A_{K,M} \right) \\ &\le Pr\left(\bigcap_{i=1}^{\frac{t}{m}} A_{K,M_i} \right) \\ &= \left[1 - \binom{k}{m} \cdot m! \cdot p^m \cdot (1-p)^{m \cdot (k-1)} \right]^{\frac{t}{m}}. \end{aligned}$$

Note that A_K is mutually independent of all the other events $A_{K'}$ except for those with $K \cap K' \ne \phi$. There are exactly

$$\binom{n}{k} - \binom{n-k}{k} - 1$$

such events. According to Corollary 2.4.3, a $t \times n$ (k, m, n)-selector exists whenever

$$e \cdot \left[1 - \binom{k}{m} \cdot m! \cdot p^m \cdot (1-p)^{m\cdot(k-1)}\right]^{\frac{t}{m}} \cdot \left[\binom{n}{k} - \binom{n-k}{k}\right] \leq 1.$$

holds. Taking natural logarithm to both sides yields the equivalent inequality

$$t \geq m \cdot \frac{1 + \ln\left[\binom{n}{k} - \binom{n-k}{k}\right]}{-\ln\left[1 - \binom{k}{m} \cdot m! \cdot p^m \cdot (1-p)^{m\cdot(k-1)}\right]}. \tag{4.15}$$

Using the fact that $-\ln(1-x) \geq x$ for $0 \leq x < 1$, we conclude that whenever

$$t \geq m \cdot \frac{1 + \ln\left[\binom{n}{k} - \binom{n-k}{k}\right]}{\binom{k}{m} \cdot m! \cdot p^m \cdot (1-p)^{m\cdot(k-1)}} \tag{4.16}$$

holds, (4.15) holds. To minimize the R.H.S. of (4.16), we let $p = \frac{1}{k}$ and complete the proof. ∎

As $m = 1$, the bound in (3.4) is $O(k \ln n)$ and the bound in (4.14) is $O((k-1)\ln n)$, which is a little bit better.

Chapter 5

Conclusion

In Theorem 4.3.2, we define a function $f_{d,r,s}(p)$ of p. To minimize the R.H.S. of the inequality (4.11), we must maximize $f_{d,r,s}(p)$, which is indeed a tough task. However, the maximum does exist, since $t(n, d, r, s]$ is a positive integer for fixed n, d, r, and s. We leave this as an open problem. Also, in our proof of Theorem 4.4.1, we partition the row indices into $\frac{t}{m}$ parts of equal size to obtain an approximation of $Pr(A_K)$. However, when $m \geq 2$, this approximation is not as good as we expect. Finally, we point out that all bounds in Chapter 4 are obtained by using probabilistic method. We do wish that deterministic constructions can be discovered in the near future.

Bibliography

[1] Noga Alon, Joel H. Spencer, The Probabilistic Method, 2nd ed., *John Wiley and Sons, Inc.*, (2000).

[2] H. B. Chen, D. Z. Du and F. K. Hwang, An unexpected meeting of four seemingly unrealated problems: graph testing, DNA complex screening, superimposed codes and secure key distribution, *J. Combin. Opt.*, to appear.

[3] H. B. Chen, H. L. Fu and F. K. Hwang, An upper bound of the number of tests in pooling designs for the error-tolerant complex model, Opt. Letters, to appear.

[4] P. Damaschke, Threshold Group Testing, *Electronic Notes in Discrete Mathematics 21*, (2005), 265-271.

[5] A. De Bonis, Leszek Gąsieniec, U. Vaccaro, Optimal Two-Stage Algorithms for Group Testing Problems, *SIAM J. Comput. Vol. 34, No. 5*, (2005), 1253-1270.

[6] A. De Bonis and U. Vaccaro, Improved algorithms for group testing with inhibitors, *Inform. Process Lett. 67*, (1998), 57-64.

[7] D. Deng, D. R. Stinson, R. Wei, The Lovász Local Lemma and Its Applications to some Combinatorial Arrays, *Designs, Codes and Cryptography, 32*, (2004), 121-134.

[8] Ding-Zhu Du, Frank K. Hwang, Combinatorial Group Testing and Its Applications, 2nd ed., *World Scientific*, (2000).

[9] Ding-Zhu Du, Frank K. Hwang, Pooling Designs and Nonadaptive Group Testing - Important Tools for DNA Sequencing, *World Scientific*, (2006).

[10] P. Erdős, L. Lovász, Problems and Results on 3-chromatic Hypergraphs and Some Related Questions, *in: Infinite and Finite Sets*(A. Hajnal et al., eds.), North-Holland, Amsterdam, (1975), 609-628.

[11] M. Farach, S. Kannan, E. Knill, S. Muthukrishnan, Group Testing Problem with Sequences in Experimental Molecular Biology, *Proc. Compression and Complexity of Sequences*, (1997), 357-367.

[12] C. H. Li, A sequential method for screening experimental variables, *J. Amer. Statist. Assoc. 57*, (1962), 455-477.

[13] L. Lovász, On the Ratio of Optimal Integral and Fractional Covers, *Discrete Math., 13*, (1975), 383-390.

[14] F. P. Ramsey, On A Problem of Formal Logic, *Proc. London Math. Soc. 30(2)*, (1929), 264-286.

[15] Joel Spencer, Probabilistic Methods, *Graphs and Combinatorics 1*, (1985), 357-382.

[16] D. R. Stinson and R. Wei, Generalized cover-free families, *Discrete Math. 279*, (2004), 463-477.

[17] Hong-Gwa Yeh, d-Disjunct matrices: bounds and Lovász Local Lemma, *Discrete Mathematics 253*, (2002), 97-107.