Pooling Spaces and Non-Adaptive Pooling Designs

Tayuan Huang* Chih-wen Weng†

October 29, 2003

Abstract

A pooling space is defined to be a ranked partially ordered set with atomic intervals. We show how to construct non-adaptive pooling designs from a pooling space. Our pooling designs are e-error detecting for some e; moreover e can be chosen to be very large compared with the maximal number of defective items. Eight new classes of non-adaptive pooling designs are given, which are related to the Hamming matroid, the attenuated space, and six classical polar spaces. We show how to construct a new pooling space from one or two given pooling spaces.

Keywords: pooling space, pooling design, ranked partially ordered set, atomic interval

1 Introduction

The basic problem of group testing is to identify the set of defective items in a large population of items. A group testing algorithm is *non-adaptive* if all tests must be specified without knowing the outcomes of other tests.

*Department of Applied Mathematics, National Chiao Tung University, Taiwan R.O.C.
†Department of Applied Mathematics, National Chiao Tung University, Taiwan R.O.C.

A non-adaptive group testing algorithm is useful in many areas. One of the examples is the problem of DNA library screening. Suppose we have n items to be tested and that there are at most d defective items among them. Each test (or pool) is (or contains) a subset of items. The output of a pool is *positive* if and only if it contains at least one of the defective items on the defective items, and the goal is to determine all of the defectives in t tests. A mathematical model of the non-adaptive group testing design for this problem is a $t \times n$ d-disjunct matrix (see Section 2). In this paper, we define a *pooling space* to be a ranked partially ordered set which has atomic intervals. We show how to construct d-disjunct matrices from a pooling space. These d-disjunct matrices have a special property described below. If we view these d-disjunct matrices as $(d-1)$-disjunct matrices, then they detect e errors for some positive integer e. As our examples show, the number e is very large compared to d. A. Macula [7], [8] gave a construction of d-disjunct matrices from the poset consisting of the subsets of a finite set. H. Ngo and D. Du [11] gave a construction of d-disjunct matrices from the poset consisting of the subspaces of a vector space. Our construction is a generalization of their results. This type of generalization was initially proposed by H. Ngo and D. Du[10, p. 177].

2 Preliminaries

Let M be a $t \times n$ matrix over $\{0, 1\}$. In this paper we frequently associate each row i (resp. column j) with a set that contains all column indices j (resp. row indices i) such that $M_{ij} = 1$. M is said to be *d-disjunct* if the union of any d columns does not contain another column. A d-disjunct $t \times n$ matrix M can be used to design a non-adaptive group testing algorithm on n items by associating the column indices with the items and the row indices with the tests. If $M_{ij} = 1$ then item j is contained in test i. Let M be a d-disjunct matrix. The *weight* $\mathrm{wt}(u)$ of a column vector or a row vector u of M is the number of 1's in u.

Example 2.1. *We can easily check*

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

is 2*-disjunct, since the union of any two columns of* M *does not contain any one of the remaining two columns. Each column of* M *has weight* 3 *and each row of* M *has weight* 2*.*

Let M be a $t \times n$ d-disjunct matrix. For a set $S \subseteq \{1, 2, \cdots, n\}$ with $|S| \leq d$, S represents the set of defective items and the *output* $o(S)$ of S in M is the union of those columns indexed by S. For example $o(\{2, 3\}) = (1, 1, 1, 0, 1, 1)^t$ with M as above (Example 2.1). Kautz and Singleton [6] gave a simple algorithm to identify the set S from its test result $u = o(S)$. In set notation, the algorithm can be written as

$$S = \{j \mid C_j \subseteq u\}, \tag{2.1}$$

where $C_1, C_2, \cdots, C_n$ are columns of M. The design of a d-disjunct matrix is also called *non-adaptive pooling design*.

A $t \times n$ matrix M over $\{0, 1\}$ is (d, e)-*disjunct* if for any $d + 1$ columns $C'_0, C'_1, \cdots, C'_d$ of M there are at least $e + 1$ elements in

$$C'_0 - \bigcup_{i=1}^{d} C'_i.$$

In particular, $(d, 0)$-disjunct is d-disjunct. In Example 2.1, M is $(2, 0)$-disjunct and $(1, 1)$-disjunct, but M is not $(2, 1)$-disjunct. From a coding theory point of view, a (d, e)-disjunct matrix is equivalent to a *superimposed distance codes* with *strength* d and *distance* $e + 1$. See [3], [4] for details.

We show that a (d, e)-disjunct matrix can be used to construct a non-adaptive pooling design that can detect e errors and correct $\lfloor \frac{e}{2} \rfloor$ errors. Let M be a (d, e)-disjunct $t \times n$ matrix. Let $S, T \subseteq \{1, 2, \cdots, n\}$ be two distinct subsets with each at most d elements. We show the Hamming distance of the test results $o(S)$ and $o(T)$ is at least $e+1$. At least one of $S-T$, $T-S$ is nonempty, so assume $S-T \neq \emptyset$. Pick $j \in S-T$. We can find $e+1$ positions i such that $M_{ij} = 1$ and $M_{ik} = 0$ for all $k \in T$. Hence $o(S)$ and $o(T)$ have Hamming distance at least $e+1$.

We now give the basic definitions and properties of a partially ordered set. The expert may want to skip the remaining of this section and go to the next section.

Let P denote a finite set. By a *partial order* on P, we mean a binary relation $\leq$ on P such that

(i) $x \leq x \quad (\forall \ x \in P)$,

(ii) $x \leq y$ and $y \leq z \quad \longrightarrow \quad x \leq z \qquad (\forall \ x, y, z \in P)$,

(iii) $x \leq y$ and $y \leq x \quad \longrightarrow \quad x = y \qquad (\forall \ x, y \in P)$.

By a *partially ordered set* (or *poset*, for short), we mean a pair $(P, \leq)$, where P is a finite set, and where $\leq$ is a partial order on P. By abusing notation, we will suppress reference to $\leq$, and just write P instead of $(P, \leq)$.

Let P denote a poset, with partial order $\leq$, and let x and y denote any elements in P. As usual, we write $x < y$ whenever $x \leq y$ and $x \neq y$. We say y *covers* x whenever $x < y$, and there is no $z \in P$ such that $x < z < y$. An element $x \in P$ is said to be *minimal* whenever there is no $y \in P$ such that $y < x$. Let $\min(P)$ denote the set of all minimal elements in P. Whenever $\min(P)$ consists of a single element, we denote it by 0, and we say P has *the least element* 0.

Throughout the paper we assume P is a poset with the least element 0. By an *atom* in P, we mean an element in P that covers 0. We let A_P denote the set of atoms in P. By a *rank function* on P, we mean a function

$$\text{rank} : P \longrightarrow \mathbb{N}$$

such that $\text{rank}(0) = 0$, and such that for all $x, y \in P$, y covers x implies $\text{rank}(y) - \text{rank}(x) = 1$. Observe the rank function is unique if it exists. P is said to be *ranked* whenever P has a rank function. In this case, we set

$$\text{rank}(P) := \max\{\text{rank}(x) | x \in P\},$$

$$P_i := \{x | x \in P, \text{rank}(x) = i\} \quad (i \in \mathbb{N} \cup \{0\}),$$

and observe $P_0 = \{0\}$, $P_1 = A_P$.

Let P denote any finite poset, and let S denote any subset of P. Then there is a unique partial order on S such that for all $x, y \in S$, $x \leq y$ in S if and only if $x \leq y$ in P. This partial order is said to be *induced* from P. By a *subposet* of P, we mean a subset of P, together with the partial order induced from P. Pick any $x, y \in P$ such that $x \leq y$. By the *interval* $[x, y]$, we mean the subposet

$$[x, y] := \{z | z \in P, x \leq z \leq y\}$$

of P.

Let P denote any poset, and let S be a subset of P. Fix $z \in P$. Then z is said to be an *upper bound* of S, if $z \geq x$ for all $x \in S$. Suppose the subposet of upper bounds of S has a unique minimal element. In this case we call this element *the least upper bound* of S.

Suppose P is ranked. Then P is said to be *atomic* whenever for each element x of P, x is the least upper bound of $[0, x] \cap P_1$.

Let q be a positive integer. Fix a positive integer N. The *Gaussian binomial coefficients with basis* q is defined by

$$\begin{bmatrix} N \\ i \end{bmatrix}_q = \begin{cases} \prod_{j=0}^{i-1} \frac{N-j}{i-j} & \text{if } q = 1, \\ \prod_{j=0}^{i-1} \frac{q^N - q^j}{q^i - q^j} & \text{if } q \neq 1. \end{cases}$$

In the case $q = 1$, for convenience, we write $\binom{N}{i}$ instead of $\begin{bmatrix} N \\ i \end{bmatrix}_1$. Now assume $q = 1$, or a prime power. Set

$$L_q(N) = \begin{cases} \text{all subsets of } \{1, 2, \cdots, N\} & \text{if } q = 1, \\ \text{subspaces of } GF(q)^N & \text{if } q \text{ is a prime power,} \end{cases}$$

where $GF(q)$ is the finite field of q elements. Let $P = L_q(N)$ be a poset with the usual set inclusion order. Note that

$$\begin{bmatrix} N \\ i \end{bmatrix}_q = |P_i|.$$

3 Construct (d, e)-disjunct Matrices

Let P be a poset. For any $w \in P$, define

$$w^+ = \{y \geq w | y \in P\}.$$

A *pooling space* is a ranked poset P such that w^+ is atomic for all $w \in P$. In particular a pooling space is atomic. If P is a pooling space, then so is w^+ for any $w \in P$. We show how to construct d-disjunct matrices from a pooling space in this section.

Theorem 3.1. *Let P be a pooling space with rank $D \geq 1$. Fix an element $x \in P_D$ and fix an integer d $(1 \leq d \leq D)$. Let $T \subseteq P_D$ be a subset such that $|T| \leq d$ and $x \notin T$. Then there exists an element $y \in [0, x] \cap P_d$ such that $y \not\leq z$ for all $z \in T$.*

Proof. We prove the theorem by induction on D. If $D = 1$ then $d = 1$ and the theorem holds by setting $y = x$. In general, pick an element $z \in T$. Then $x \neq z$ by assumption. Since x is the least upper bound of $[0, x] \cap P_1$ and $x \not\leq z$, z is not an upper bound of $[0, x] \cap P_1$. Hence we can pick an element $w \in [0, x] \cap P_1$ such that $w \not\leq z$. Then $T \cap w^+$ has at most $d - 1$ elements. In the pooling space w^+, the element x and the elements of $T \cap w^+$ all have rank $D-1$, and the elements of $w^+ \cap P_d$ have rank $d-1$. Hence by induction, we can choose $y \in [w, x] \cap P_d$ such that $y \not\leq u$ for all $u \in T \cap w^+$. Note that clearly $y \not\leq u$ for all $u \in T \setminus w^+$. This proves the theorem. □

With notation in Theorem 3.1, observe for any integer ℓ $(d \leq \ell \leq D)$, each element $w \in [y, x] \cap P_\ell$ satisfies $w \leq x$ and $w \not\leq z$ for all $z \in T$. Hence the characteristic matrix of the binary relation induced on the subposet $P_\ell \cup P_D$ of a pooling space P is in fact (d, e)-disjunct, where the number $e + 1$ is the minimal number in counting such w. More precisely, we state this as the following corollary.

Corollary 3.2. *Let P be a pooling space with rank D. Fix an integer ℓ $(1 \leq \ell \leq D)$. Let $M = M(D, \ell)$ be the matrix over $\{0, 1\}$ whose rows (resp. columns) are indexed by P_ℓ (resp. P_D) such that $M_{uv} = 1$ iff $u \leq v$. Then for each integer d $(1 \leq d \leq \ell)$, M is (d, e)-disjunct, where*

$$e = \min |\bigcup [y, x] \cap P_\ell| - 1$$

with the minimum taken over all pairs (x, T) *such that* $x \in P_D$, $T \subseteq P_D$, $x \notin T$, $|T| \leq d$, *and with the union taken over all* y *such that* $y \in P_d$, $y \leq x$, $y \not\leq z$ *for all* $z \in T$.

Note that the *truncation* of a pooling space is a pooling space. That is if P is a pooling space with rank D, then

$$P_0 \cup P_1 \cup \cdots \cup P_k$$

is a pooling space with rank k for each k $(0 \leq k \leq D)$. Hence in the above construction of M we can choose any k $(\ell \leq k \leq D)$ and use P_k to replace P_D. The definition of e in Corollary 3.2 seems complicate. However, in our examples in the next section the number $|[y, x] \cap P_\ell|$ is a constant.

4 Examples

In this section, we give some examples of pooling spaces P with rank D. All of these examples are *quantum matroids* with the base q [13], where q is 1 or a prime power. The number $|P_i|$ can be computed from results given in [13]. We omit the details of the computing. For integers $1 \leq d \leq \ell \leq k \leq D$, the examples produce the (d, e)-disjunct matrices $M = M(k, \ell)$ have size $t \times n$, where $t = |P_\ell|$, $n = |P_k|$ and

$$e = \begin{bmatrix} k-d \\ \ell-d \end{bmatrix}_q - 1.$$

The weight of each column of M is

$$\begin{bmatrix} k \\ \ell \end{bmatrix}_q,$$

and the weight of each row of M is

$$\frac{|P_k|}{|P_\ell|} \begin{bmatrix} k \\ \ell \end{bmatrix}_q.$$

1. The Hamming matroid $H(D, N)$ $(2 \leq N)$ [2], [12].

Set

$$A = A_1 \cup A_2 \cup \cdots \cup A_D \qquad \text{(disjoint union)},$$

where

$$|A_i| = N \qquad (1 \le i \le D).$$

$P = \{x \mid x \subseteq A, |x \cap A_i| \le 1 \ \text{ for } \ \text{all } \ i \quad (1 \le i \le D)\}$,

$x \le y$ whenever x is a subset of $y \quad (x, y \in P)$,

$\text{rank}(x) = |x| \quad (x \in P)$,

$|P_i| = \binom{D}{i} N^i$.

In [9], A. J. Macula and P. A. Vilenkin implicitly gave this construction too.

2. The attenuated space $A_q(D, N)$ $(D \le N)$ [2], [5].

Let V denote a vector space of dimension N over the field $GF(q)$, and fix a subspace $w \subseteq V$ of dimension $N - D$.

$P = \{x \mid x \ \text{ is a subspace of } \ V, \quad x \cap w = 0\}$,

$x \le y$ whenever x is a subspace of $y \quad (x, y \in P)$,

$\text{rank}(x) = \dim(x) \quad (x \in P)$,

$|P_i| = \begin{bmatrix} D \\ i \end{bmatrix}_q q^{i(N-D)}$.

3. The classical polar spaces of rank D over $GF(q)$ [1].

Let V denote a vector space over the field $GF(q)$, and assume V possesses a given non-degenerate form. We call a subspace of V *isotropic* whenever the form vanishes completely on that subspace. The maximal isotropic subspaces have the same dimension, denoted by D.

$P = \{x \mid x \text{ is an isotropic subspace of } V\}$,

$x \leq y$ whenever x is a subspace of y $\quad (x, y \in P)$,

$\mathrm{rank}(x) = \dim(x) \quad (x \in P)$,

name	**dim** V	**form**	$\lvert P_i \rvert$
$B_D(q)$	$2D+1$	quadratic	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^D)(1+q^{D-1})\cdots(1+q^{D-i+1})$
$C_D(q)$	$2D$	alternating	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^D)(1+q^{D-1})\cdots(1+q^{D-i+1})$
$D_D(q)$	$2D$	quadratic (Witt index D)	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^{D-1})(1+q^{D-2})\cdots(1+q^{D-i})$
${}^2D_{D+1}(q)$	$2D+2$	quadratic (Witt index D)	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^{D+1})(1+q^{D})\cdots(1+q^{D-i+2})$
${}^2A_{2D}(r)$	$2D+1$	Hermitian ($q = r^2$)	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^{D+\frac{1}{2}})(1+q^{D-\frac{1}{2}})\cdots(1+q^{D-i+\frac{3}{2}})$
${}^2A_{2D-1}(r)$	$2D$	Hermitian ($q = r^2$)	$\begin{bmatrix} D \\ i \end{bmatrix}_q (1+q^{D-\frac{1}{2}})(1+q^{D-\frac{3}{2}})\cdots(1+q^{D-i+\frac{1}{2}})$

5 Pooling Polynomials

Let P be a pooling space with rank D. The ratio $\frac{|P_\ell|}{|P_k|}$ is the main concern of the construction of pooling designs, and the structure of P is less important. With this motivation, we give the following definition.

Definition 5.1. Let P be a pooling space with rank D. The *pooling polynomial* of P is

$$f_P(x) := \sum_{i=0}^{D} |P_i| x^i.$$

Note that the constant term of a pooling polynomial is always 1. With lexicographical order, 1 and $1+x$ are the first two pooling polynomials.

Let $P^{'}$, $P^{''}$ be pooling spaces with rank $D^{'}$, $D^{''}$ respectively. We define the *direct sum* $P^{'}+P^{''}$ of $P^{'}$ and $P^{''}$ as follows. The element set of $P^{'}+P^{''}$ is the disjoint union of $P^{'}$ and $P^{''}$ except that the 0 of $P^{'}$ and the 0 of $P^{''}$ are identical. Hence $P^{'}+P^{''}$ has $|P^{'}|+|P^{''}|-1$ elements. The partial order of $P^{'}+P^{''}$ is naturally inherited from $P^{'}$ and $P^{''}$. It is easy to see $P^{'}+P^{''}$ is a pooling space with rank $\max\{D^{'},D^{''}\}$. We define the *product* $P^{'}\bigotimes P^{''}$ of $P^{'}$ and $P^{''}$ as follows. The element set of $P=P^{'}\bigotimes P^{''}$ is

$$\{(a,b)\mid a\in P^{'},b\in P^{''}\}.$$

The partial order in $P^{'}\bigotimes P^{''}$ is defined by

$$(a,b)\leq(c,d) \qquad \text{iff} \qquad a\leq c \text{ and } b\leq d,$$

for any $a,c\in P^{'}$ and any $b,d\in P^{''}$. It is easy to see that for any $a,c\in P^{'}$ and $b,d\in P^{''}$, the following (i)-(iii) hold.

(i) $\operatorname{rank}((a,b))=\operatorname{rank}(a)+\operatorname{rank}(b)$;

(ii) $[0,(a,b)]\cap P_1=\{(a_1,0),\cdots,(a_r,0),(0,b_1),\cdots,(0,b_s)\}$, where $\{a_1,\cdots,a_r\}=[0,a]\cap P_1^{'}$ and $\{b_1,\cdots,b_s\}=[0,b]\cap P_1^{''}$.

(iii) $[(a,b),(c,d)]=[a,c]\bigotimes[b,d]$.

We conclude from (i)-(iii) above that $P^{'}\bigotimes P^{''}$ is a pooling space with rank $D^{'}+D^{''}$.

Note that if P is a pooling space then so is $P\backslash w^{+}$ for any $w\in P$. Let f be a pooling polynomial. By a *reduction* of f, we mean a polynomial obtained by replacing the leading coefficient of f by a smaller nonnegative integer. We immediately have the following theorem.

Theorem 5.2. *Let $\mathcal{F}$ be the set of pooling polynomials. Suppose $f_1(x),f_2(x)\in\mathcal{F}$. Then the following (i)-(iii) hold.*

(i) A reduction of $f_1(x)$ is in $\mathcal{F}$;

(ii) $f_1(x)+f_2(x)-1\in\mathcal{F}$;

(iii) $f_1(x)f_2(x) \in \mathcal{F}$.

Theorem 5.2 provides us a few ways to construct more pooling polynomials and corresponding pooling designs.

Example 5.3. $(1+3x+2x^2)^m$ *is a pooling polynomial, since it can be obtained from the pooling polynomial* $1+x$ *by using productions and reductions as shown in the equation*

$$(1+3x+2x^2)^m = (((1+x)^3 - x^3) - x^2)^m.$$

6 Concluding Remarks

We construct (d, e)-disjunct matrices from a pooling space in Section 3. Some examples of pooling spaces are given in Section 4. By checking these examples, the ratio $\frac{t}{n} = \frac{|P_\ell|}{|P_k|}$ is small and the error-tolerance number e is large if ℓ, k are well chosen. However it seems that d is too small compared to n in all these examples. We show how to construct a new pooling space from given pooling spaces in Section 5. This can be used to obtain a pooling space with a desired $|P_i|$ range.

Of course, our list of pooling spaces is not exhaustive. It can be expected that there are a lot of unknown pooling spaces and a complete list of them is unlikely to be completed. We give another class to show this line of study might have number theory involved. Fix a positive integer m, and set

$$P = \{i \mid 2 \leq i \leq m, \text{ and } i \text{ is an integer which contains no square factors}\}.$$

The partial order in P is defined by

$$i \leq j \qquad \text{iff} \qquad i \text{ divides } j.$$

By identifying an element in P with a subset of primes, the poset P can be obtained from the infinite poset consisting all the subsets of primes and then deleting each subposet w^+ for each integer $w > m$ (in natural integers ordering). It can be easily checked that P is a pooling space. However the computing of $|P_i|$ is not likely to be written as a nice formula of i and m.

Another interesting problem is to find an effective decoding algorithm for the set $S \subseteq \{1, 2, \cdots, n\}$ of defective items from its output u with at most

$\lfloor \frac{e}{2} \rfloor$ errors in a (d, e)-disjunct matrix M. This will be a generalization of the well known decoding algorithm in the d-disjunct case. See [6] for details.

A class of pooling space related to the Hermitian form graphs is constructed in [14]. All examples of the pooling spaces we mentioned in this paper have an additional property of being *(meet) semi-lattice*; this means that any two elements have a greatest lower bound. To close the paper, we propose the following question. Try to find a pooling space which is not a semi-lattice.

Acknowledgements The authors wish to express their sincere thanks to Frank Hwang and anonymous referees for very helpful comments and suggestions that led to a considerable improvement of this paper.

References

[1] P. Cameron, Projective and polar spaces, QMW Maths Notes 13.

[2] P. Delsarte, Association schemes and t-designs in regular semi-lattice, J. Combin. Theory Ser. A, 20:230-243, 1976.

[3] A. G. D'yachkov and V. Rykov, Superimposed distance codes, Problems Control and Information Theory, 18(4):237-250, 1989.

[4] A. G. D'yachkov, A.J. Macula and P. A. Vilenkin, Nonadaptive Group Testing with Error-Correction d^e-Disjunct Inclusion Matrices, preprint.

[5] T. Huang, A characterization of the association schemes of bilinear forms, European J. Combin., 8:159-173, 1987.

[6] W. H. Kautz and R. R. Singleton, Nonadaptive binary superimposed codes, IEEE Trans. Inform. Theory, 10: 363-377, 1964.

[7] A. J. Macula, A simple construction of d-disjunct matrices with certain constant weights, Discrete Math., 162:311-312, 1996.

[8] A. J. Macula, Probabilistic nonadaptive group testing in the presence of errors and dna library screening, Annals of Combinatorics, 3:61-69, 1999.

[9] A. J. Macula and P. A. Vilenkin, Constructions of superimposed codes based on incidence structures, IEEE ISIT, Sorrento, Italy, June 25-30, 2000.

[10] H. Ngo and D. Zu, A Survey on Combinatorial Group Testing Algorithms with Applications to DNA Library Screening, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 55:171-182, 2000.

[11] H. Ngo and D. Du, New Constructions of Non-Adaptive and Error-Tolerance Pooling Designs, Discrete Math., 243:161-170, 2002.

[12] P. Terwilliger, The incidence algebra of a uniform poset, Coding Theory and Design Theory part I: Coding Theory, IMA volumes in Mathematics and its applications Vol. 20, Springer, New York, 193-212, 1990.

[13] P. Terwilliger, Quantum matroids, Progress in algebraic combinatorics (Fukuoka, 1993), 323–441, Adv. Stud. Pure Math., 24, Math. Soc. Japan, Tokyo, 1996.

[14] C. Weng, D-bounded Distance-regular Graphs, Europ. J. Combinatorics, 18:211-229, 1997.