# Constructing error-correcting pooling designs with singular linear space

Xuemei Liu *, You Gao

*College of Science, Civil Aviation University of China, Tianjin, 300300, P.R.China*

**Abstract** Pooling designs are standard experimental tools in many biotechnical applications. In this paper, we construct a family of error-correcting pooling designs with the incidence matrix of two types of subspaces of singular linear space over finite fields, and exhibit their disjunct properties.

***Keywords:*** Pooling designs, $d^e$-disjunct matrix, Singular linear space

***AMS** classification*: 20G40 51D25

## 1. Introduction

Given a set of n items with some defectives, the group testing problem is asking to identify all defectives with the minimum number of tests each of which is on a subset of items, called a pool, and the test-outcome is negative if the pool does not contain any defective and positive if the pool contains a defective. A pooling design is a group testing algorithm of special type, also called nonadaptive group testing, in which all pools are given at the beginning of the algorithm so that no test-outcome of one pool can effect the determination of another pool. Objectives of group testing vary from minimizing the number of tests, limiting number of pools, limiting pool sizes to tolerating a few errors. It is conceivable that these objectives are often contradicting, thus testing strategies are application dependent.

A pooling design can be represented by a binary matrix whose columns are indexed with items and rows are indexed with pools; an entry at cell $(i, j)$ is 1 if the $i$-th pool contains the $j$-th item, and 0, otherwise. Such a binary matrix is said to be $d$-disjunct. With $d$-disjunct pooling design, decoding is very simple. Remove all items in negative pools. The remaining items are all defectives. In practice, test-outcomes may contain errors. To make pooling design error tolerant, one introduces the concept of $d^e$-disjunct matrix (see Macula [1]). A binary matrix $M$ is said to be $d^e$-disjunct if given any $d+1$ columns of $M$ with one designated, there are

*Correspondence : College of Science, Civil Aviation University of China,Tianjin,300300, P.R.China; E-mail: xm-liu771216@163.com.

$e+1$ rows with a 1 in the designated column and 0 in each of the other $d$ columns. The $d^0$-disjunctness is actually the $d$-disjunctness. $D'$yachkov et al. (see [2]) proposed the concept of fully $d^e$-disjunct matrices. An $d^e$-disjunct matrix is fully $d^e$-disjunct if it is neither $(d+1)^e$- disjunct nor $d^{e+1}$-disjunct.

The Pooling designs have many applications in molecular biology, such as DNA library screening, nonunique probe selection, gene detection, etc. (Du and Hwang [3]; Du et al. [4]; D yachkov et al. [5]). There are several constructions of $d^e$-disjunct matrices in the literature (Guo et al. [6]; Balding and Torney [7]; Erdös et al. [8]; Guo [9]; Huang and Weng [10]; Li et al. [11]; Macula [12]; Nan and Guo [13]; Ngo and Du [14]; Zhang et al. [15],[16]). In this paper we present a new construction associated with subspaces in $\mathbb{F}_q^{(n+l)}$, and exhibit their disjunct properties.

## 2.Singular linear spaces

In this section we shall introduce the concepts of subspaces of type $(m,k)$ in singular linear space.(see Wang et al. [17])

Let $\mathbb{F}_q$ be a finite field with $q$ elements, where $q$ is a prime power. For two non-negative integers $n$ and $l$, $\mathbb{F}_q^{(n+l)}$ denotes the $(n+l)$-dimensional row vector space over $\mathbb{F}_q$. The set of all $(n+l)\times(n+l)$ nonsingular matrices over $\mathbb{F}_q$ of the form

$$\begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix},$$

where $T_{11}$ and $T_{22}$ are nonsingular $n\times n$ and $l\times l$ matrices, respectively, forms a group under matrix multiplication, called the singular general linear group of degree $n+l$ over $\mathbb{F}_q$ and denoted by $GL_{n+l,n}(\mathbb{F}_q)$. If $l=0$(resp. $n=0$), $GL_{n,n}(\mathbb{F}_q)=GL_n(\mathbb{F}_q)$ (resp. $GL_{l,0}(\mathbb{F}_q)=GL_l(\mathbb{F}_q)$) is the general linear group of degree $n$ (resp. $l$). (See Wan [18]) Let $P$ be an $m$-dimensional subspace of $\mathbb{F}_q^{(n+l)}$ , denote also by $P$ an $m\times(n+l)$ matrix of rank $m$ whose rows span the subspace $P$ and call the matrix $P$ a matrix representation of the subspace $P$. There is an action of $GL_{n+l,n}(\mathbb{F}_q)$ on $\mathbb{F}_q^{(n+l)}$ defined as follows

$$\mathbb{F}_q^{(n+l)}\times GL_{n+l,n}(\mathbb{F}_q)\to\mathbb{F}_q^{(n+l)}$$
$$((x_1,\ldots,x_n,x_{n+1},\ldots,x_{n+l}),T)\mapsto (x_1,\ldots,x_n,x_{n+1},\ldots,x_{n+l})T$$

The above action induces an action on the set of subspaces of $\mathbb{F}_q^{(n+l)}$, i.e., a subspace $P$ is carried by $T\in GL_{n+l,n}(\mathbb{F}_q)$ to the subspace $PT$. The vector space $\mathbb{F}_q^{(n+l)}$ together with the above group action, is called the $(n+l)$-dimensional singular linear space over $\mathbb{F}_q$. For $1\le i\le n+l$, let $e_i$ be the row vector in $\mathbb{F}_q^{(n+l)}$ whose $i$-th coordinate is 1 and all other coordinates are 0. Denote by $E$ the $l$-dimensional subspace of $\mathbb{F}_q^{(n+l)}$ generated by $e_{n+1}, e_{n+2}, \ldots, e_{n+l}$. An $m$-dimensional subspace P of $\mathbb{F}_q^{(n+l)}$ is called a subspace of type $(m,k)$ if $\dim(P \cap E) = k$. The collection of all the subspaces of types $(m,0)$ in $\mathbb{F}_q^{(n+l)}$, where $0 \le m \le n$, is the attenuated space.(see A.E. Brouwer et al. [19])

Let $\mathcal{M}(m,k;n+l,n)$ denote the set of all the subspaces of type $(m,k)$ of $\mathbb{F}_q^{(n+l)}$. By Wang et al. [20], $\mathcal{M}(m,k;n+l,n)$ forms an orbit under the action of $GL_{n+l,n}(\mathbb{F}_q)$.

We begin with some useful propositions.

Let $\mathcal{M}(m,k;n+l,n)$ denote the set of all subspaces of type $(m,k)$ in $\mathbb{F}_q^{(n+l)}$, and let $N(m,k;n+l,n)$ denote the size of $\mathcal{M}(m,k;n+l,n)$.

**Proposition 2.1** (Wang et al. [17] Lemma 2.1). $\mathcal{M}(m,k;n+l,n)$ is non-empty if and only if $0 \le k \le l$ and $0 \le m-k \le n$. Moreover, if $\mathcal{M}(m,k;n+l,n)$ is non-empty, then it forms an orbit of subspaces under $GL_{n+l,n}(\mathbb{F}_q)$ and

$$N(m,k;n+l,n) = q^{(m-k)(l-k)} \begin{bmatrix} n \\ m-k \end{bmatrix}_q \begin{bmatrix} l \\ k \end{bmatrix}_q .$$

For a fixed subspace $P$ of type $(m,k)$ in $\mathbb{F}_q^{(n+l)}$, let $\mathcal{M}(m_1,k_1;m,k;n+l,n)$ denote the set of all the subspaces of type $(m_1,k_1)$ contained in $P$, and let $N(m_1,k_1;m,k;n+l,n) = |\mathcal{M}(m_1,k_1;m,k;n+l,n)|$.

By the transitivity of $GL_{n+l,n}(\mathbb{F}_q)$ on the set of subspaces of the same type, $N(m_1,k_1;m,$
$k;n+l,n)$ is independent of the particular choice of the subspace $P$ of type $(m,k)$.

**Proposition 2.2.**(Wang et al. [17] Lemma 2.2) $\mathcal{M}(m_1,k_1;m,k;n+l,n)$ is non-empty if and only if $0 \le k_1 \le k \le l$ and $0 \le m_1 - k_1 \le m-k \le n$. Moreover, if $\mathcal{M}(m_1,k_1;m,k;n+l,n)$ is non-empty, then

$$N(m_1,k_1;m,k;n+l,n) = q^{(m_1-k_1)(k-k_1)} \begin{bmatrix} m-k \\ m_1-k_1 \end{bmatrix}_q \begin{bmatrix} k \\ k_1 \end{bmatrix}_q .$$

For a fixed subspace $P$ of type $(m_1,k_1)$ in $\mathbb{F}_q^{(n+l)}$, let $\mathcal{M}'(m_1,k_1;m,k;n+l,n)$ denote the set of all the subspaces of type $(m,k)$ containing $P$, and let $N'(m_1,k_1;m,k;n+l,n) = |\mathcal{M}'(m_1,k_1;m,k;n+l,n)|$. By the transitivity of $GL_{n+l,n}(\mathbb{F}_q)$ on the set of subspaces of the same type, $N'(m_1,k_1;m,k;n+l,n)$ is independent of the particular choice of the subspace $P$ of type

$(m_1, k_1)$.

**Proposition 2.3.**(Wang et al. [17] Lemma 2.3) $\mathcal{M}'(m_1, k_1; m, k; n+l, n)$ is non-empty if and only if $0 \le k_1 \le k \le l$ and $0 \le m_1 - k_1 \le m - k \le n$. Moreover, if $\mathcal{M}'(m_1, k_1; m, k; n+l, n)$ is non-empty, then

$$N'(m_1, k_1; m, k; n+l, n) = q^{(l-k)(m-k-m_1+k_1)} \begin{bmatrix} n-(m_1-k_1) \\ (m-k)-(m_1-k_1) \end{bmatrix}_q \begin{bmatrix} l-k_1 \\ k-k_1 \end{bmatrix}_q.$$

**Proposition 2.4.**(Wan [18] Corollary1.9) Let $0 \le k \le m \le n$. Then the number $N'(k, m, n)$ of $m$-dimensional vector subspaces containing a given $k$-dimensional vector subspace $\mathbb{F}_q^{(n)}$ is equal to $\begin{bmatrix} n-k \\ m-k \end{bmatrix}_q$.

**Theorem 2.5.** Given integers $0 \le k \le l$ and $0 \le m-k \le n$, the sequence $N(m, k; n+l, n)$ is unimodal and gets its peak at $m = \lfloor \frac{n+l+k}{2} \rfloor$.

**Proof** By Proposition 2.1, if $m_1 < m_2$ then we have

$$\frac{N(m_1, k; n+l, n)}{N(m_2, k; n+l, n)} = \frac{q^{(m_1-k)(l-k)} \begin{bmatrix} n \\ m_1-k \end{bmatrix}_q \begin{bmatrix} l \\ k \end{bmatrix}_q}{q^{(m_2-k)(l-k)} \begin{bmatrix} n \\ m_2-k \end{bmatrix}_q \begin{bmatrix} l \\ k \end{bmatrix}_q}$$

$$= \frac{1}{q^{(m_2-m_1)(l-k)}} \cdot \frac{(q^{m_2-k}-1)(q^{m_2-k-1}-1)\cdots(q^{m_1-k+1}-1)}{(q^{n-m_1+k}-1)(q^{n-m_1+k-1}-1)\cdots(q^{n-m_2+k+1}-1)}$$

$$= \frac{(q^{m_2-k}-1)(q^{m_2-k-1}-1)\cdots(q^{m_1-k+1}-1)}{(q^{n+l-m_1}-q^{l-k})(q^{n+l-m_1-1}-q^{l-k})\cdots(q^{n+l-m_2+1}-q^{l-k})}$$

$$= \frac{q^{m_1-k+1}-1}{q^{n+l-m_1}-q^{l-k}} \cdot \frac{q^{m_1-k+2}-1}{q^{n+l-m_1-1}-q^{l-k}} \cdot \cdots \cdot \frac{q^{m_2-k}-1}{q^{n+l-m_2+1}-q^{l-k}},$$

where $\frac{q^{m_1-k+1}-1}{q^{n+l-m_1}-q^{l-k}} < \frac{q^{m_1-k+2}-1}{q^{n+l-m_1-1}-q^{l-k}} < \cdots < \frac{q^{m_2-k}-1}{q^{n+l-m_2+1}-q^{l-k}}$.

If $m_2 \le \lfloor \frac{n+l+k}{2} \rfloor$, then $m_2 - k < n + l - m_2 + 1$, $\frac{q^{m_2-k}-1}{q^{n+l-m_2+1}-q^{l-k}} < 1$. Hence, when $0 < m_1 < m_2 \le \lfloor \frac{n+l+k}{2} \rfloor$, we have $\frac{N(m_1,k;n+l,n)}{N(m_2,k;n+l,n)} < 1$.

If $m_1 \ge \lfloor \frac{n+l+k}{2} \rfloor$, then $m_1 - k + 1 > n + l - m_1$, $\frac{q^{m_1-k+1}-1}{q^{n+l-m_1}-q^{l-k}} > 1$. Hence, when $\lfloor \frac{n+l+k}{2} \rfloor \le m_1 < m_2 < n+l$, we have $\frac{N(m_1,k;n+l,n)}{N(m_2,k;n+l,n)} > 1$. □

## 3. The construction

In this section, we construct a family of inclusion matrices associated with subspaces of $\mathbb{F}_q^{(n+l)}$, and exhibit its disjunct property.

**Definition 3.1** Given integers $0 \le k \le l, 0 \le m-k \le n, 0 \le r \le m-k-2$. Let $M(r;m,k;n+l,n)$ be the binary matrix whose rows (resp. columns) are indexed by $\mathcal{M}(r,0;n+l,n)$ (resp. $\mathcal{M}(m,k;n+l,n)$. We also order elements of these sets lexicographically. $M(r;m,k;n+l,n)$ has a 1 in row $i$ and column $j$ if and only if the $i$-th subspace of $\mathcal{M}(r,0;n+l,n)$ is a subspace of the $j$-th subspace of $\mathcal{M}(m,k;n+l,n)$.

By Propositions 2.1, 2.2 and 2.3, $M(r;m,k;n+l,n)$ is a $N(r,0;n+l,n) \times N(m,k;n+l,n)$ matrix, whose constant row (resp. column) weight is $N'(r,0;m,k;n+l,n)$ (resp. $N(r,0;m,k;n+l,n)$). Theorem 2.5 tells us how to choose $m$ so that the test to item is minimized.

**Theorem 3.2** Given $2 \le k < l, 0 \le m-k \le n, 0 \le r \le m-k-2$ and let $t = N(r,0;m,k;n+l,n)$, $u = N(r,0;m-1,k;n+l,n))$, $v = N(r,0;m-1,k-1;n+l,n)$, $x = N(r,0;m-2,k;n+l,n)$, $y = N(r,0;m-2,k-1;n+l,n)$, $z = N(r,0;m-2,k-2;n+l,n)$ and $w = \max\{u-x, u-y, u-z, v-x, v-y, v-z\}$, if $1 \le d \le \lfloor \frac{t-\max\{u,v\}-1}{w} \rfloor + 1$ then $M(r;m,k;n+l,n)$ is $d^e$-disjunct, where $e = t - \max\{u,v\} - (d-1)w - 1$. In particular, if $1 \le d \le \min\{\lfloor \frac{t-\max\{u,v\}-1}{w} \rfloor + 1, q+1\}$, then $M(r;m,k;n+l,n)$ is fully $d^e$-disjunct.

**Proof** Let $P, P_1, P_2, \cdots, P_d$ be $d+1$ distinct columns of $M(r;m,k;n+l,n)$. To obtain the maximum numbers of subspaces of $\mathcal{M}(r,0;n+l,n)$ in

$$P \cap \bigcup_{i=1}^{d} P_i = \bigcup_{i=1}^{d} (P \cap P_i),$$

we may assume that $\dim(P \cap P_i) = m-1$ and $\dim(P \cap P_i \cap P_j) = \dim((P \cap P_i) \cap (P \cap P_j)) = m-2$, for any two distinct $i$ and $j$ , where $1 \le i,j \le d$. Since $P \in \mathcal{M}(m,k;n+l,n)$, $P \cap P_i$ (resp. $P \cap P_i \cap P_j$) is a subspace of type $(m-1,k)$ or type $(m-1,k-1)$(resp. type $(m-2,k)$ or type $(m-2,k-1)$ or type $(m-2,k-2)$) of $\mathbb{F}_q^{(n+l)}$ by Proposition 2.1. By Proposition 2.2, $x > 0$, $y > 0$ and $z > 0$. By Proposition 2.2, the number of subspaces of $P$ not covered by $P_1, P_2, \cdots, P_d$ is at least

$$\begin{aligned} & t - \max\{u,v\} - (d-1)[\max\{u,v\} - \min\{x,y,z\}] \\ = {} & t - d\max\{u,v\} + (d-1) \times \min\{x,y,z\} \\ = {} & t - \max\{u,v\} - (d-1)w. \end{aligned}$$

Hence, we may take $e = t - \max\{u, v\} - (d-1)w - 1$ under the assumption that $d$. Since $e \geq 0$, we obtain

$$d \leq \lfloor \frac{t - \max\{u, v\} - 1}{w} \rfloor + 1.$$

Now we show that the maximal dimension of $P \cap \bigcup_{i=1}^{d} P_i$ is achieved by an explicit construction. For $P \cap P_1$, by Proposition 2.2, $N(r, 0; m-2, k; n+l, n) \geq 1$ , $N(r, 0; m-2, k-1; n+l, n) \geq 1$ and $N(r, 0; m-2, k-2; n+l, n) \geq 1$. Hence there exists an $(m-2)$-dimensional subspace contained in $P \cap P_1$, denoted by $Q$, such that the number of subspace of type $(r, 0)$ contained in $Q$ is equal to $\min\{x, y, z\}$. By Proposition 2.4, the number of $(m-1)$-dimensional subspaces containing $Q$ and contained in $P$ is equal to $q+1$, and each of these subspaces is a subspace of type $(m-1, k)$ or type $(m-1, k-1)$. For $1 \leq d \leq \min\{\lfloor \frac{t-\max\{u,v\}-1}{w} \rfloor + 1, q+1\}$, we choose $d$ distinct $(m-1)$-dimensional subspaces between $Q$ and $P$, say $Q_i, (1 \leq i \leq d)$. Since $N'(m-1, k; m, k; n+l, n) \geq 2$ and $N'(m-1, k-1; m, k; n+l, n) \geq 2$ by Proposition 2.3, for each $Q_i$, we can choose a subspace of type $(m, k)$ denoted by $P_i$, such that $P \cap P_i = Q_i$. Hence, each pair of $P_i$ and $P_j$ overlaps at the same subspace $Q$.

Now we have showed that $M(r; m, k; n+l, n)$ is $d^e$-disjunct. Moreover, by the assumption of $e$, we have that $M(r; m, k; n+l, n)$ is $d^e$-disjunct but not $d^{e+1}$-disjunct. On the other hand we assume that $M(r; m, k; n+l, n)$ is $(d+1)^{e'}$-disjunct. By the maximality of $e$, we infer that $e' \leq e = y - z - (d+1-1)x - 1t - \max\{u, v\} - (d+1-1)w - 1 < t - \max\{u, v\} - (d-1)w - 1 = e$. Hence $M(r; m, k; n+l, n)$ is not $(d+1)^e$-disjunct. Therefore, $M(r; m, k; n+l, n)$ is fully $d^e$-disjunct. This completes the proof. □

**Acknowledgements** This work is supported by the National Natural Science Foundation of China under Grant No. 61179026 and Supported by the Fundamental Research Funds for the Central Universities (3122013k001).

# References

[1] Macula AJ (1996) A simple construction of $d$–disjunct matrices with certain constant weights. Discrete Math 162 : 311 – 312.

[2] D'yachkov AG, Macula AJ, Vilenkin PA (2007) Nonadaptive group and trivial two-stage group testing with error-correction $d^e$–disjunct inclusion matrices. In: Csiszár I, Katona GOH, Tardos G (eds)

Entropy,search, complexity,1st edn.Springer, Berlin, pp 71-84.ISBN-10:3540325735;ISBN-13:978-3540325734.

[3] Du D, Hwang F (2006) Pooling designs and non−adaptive group testing: important tools for DNA sequencing. World Scientific,Singapore.

[4] Du D, Hwang F, Wei W, Znati T (2006) New construction for transversal design. J Comput Biol 13 : 990 − 995.

[5] D'yachkov AG, Hwang FK, Macula AJ, Vilenkin PA, Weng C (2005) A construction of pooling designs with some happy surprises. J Comput Biol 12 : 1129 − 1136.

[6] Guo J,Wang Y,Gao S,Yu J,Wu W,(2010) Constructing error-correcting pooling designs with symplectic space J comb Optim 20 : 413 − 421.

[7] Balding DJ, Torney DC (1996) Optimal pooling designs with error detection. J Comb Theory Ser A 74 : 131 − 140.

[8] Erdös P, Frankl P, Füredi D (1985) Families of finite sets in which no set is covered by the union of $r$ others. Isr J Math 51 : 79 − 89.

[9] Guo J (2010) Pooling designs associated with unitary space and ratio efficiency comparison. J Comb Optim 19 : 492 − 500.

[10] Huang T, Weng C (2004) Pooling spaces and non−adaptive pooling designs. Discrete Math 282 : 163 − 169.

[11] Li Z, Gao S, Du H, Zou F, Wu W (2010) Two constructions of new error−correcting pooling designs from orthogonal spaces over a finite field of characteristic 2. J Comb Optim 20 : 325 − 334.

[12] Macula AJ (1997) Error-correcting non-adaptive group testing with de-disjunct matrices. Discrete Appl Math 80 : 217 − 222.

[13] Nan J, Guo J (2010) New error-correcting pooling designs associated with finite vector spaces. J Comb Optim 20 : 96 − 100.

[14] Ngo H, Du D (2002) New constructions of non−adaptive and error-tolerance pooling designs. Discrete Math 243 : 167 − 170.

[15] Zhang G, Li B, Sun X, Li F (2008) A construction of $d^z$-disjunct matrices in a dual space of symplectic space. Discrete Appl Math 28 : 2400 − 2406.

[16] Zhang X, Guo J, Gao S (2009) Two new error-correcting pooling designs from $d$−bounded distance−regular graphs. J Comb Optim 17 : 339 − 345.

[17] Wang K, Guo J, Li F (2011) Singular linear space and its applications. Finite Fields Appl 17 : 395 – 406.

[18] Wan Z (2002) Geometry of classical groups over finite fields, 2nd edn. Science, Beijing.

[19] A.E. Brouwer, A.M. Cohen, A. Neumaier (1989) Distance-Regular Graphs, Springer-Verlag, Berlin, Heidelberg.

[20] K. Wang, J. Guo, F. Li (2010) Association schemes based on attenuated spaces, European J. Combin.31 297 – 305.