[*] This research was supported in part by the U.S. Air Force Office of Scientific Research under Contract No. AFOSR-68-1415.

[1] Ph.D. dissertation written under the direction of Professor R. C. Bose.

# COMBINATORIAL INFORMATION RETRIEVAL SCHEMES[*,1]

N. Vijayaditya

*Department of Statistics*
*University of North Carolina at Chapel Hill*

Institute of Statistics Mimeo Series No. 833

July, 1972

N. VIJAYADITYA. Combinatorial Information Retrieval Schemes
(Under the direction of R. C. BOSE.)

The development of an information filing scheme deals not only with the storage of data but also with the retrieval. The efficiency of a filing scheme is measured not only in terms of the ease with which it is possible to retrieve information pertinent to a given task but also in terms of the retrieval time. In this research, combinatorial methods are applied to obtain filing schemes that are efficient in terms of retrieval time. In the beginning, a brief review of the relevant basic combinatorial techniques along with applications to the filing schemes is given. Then a few interesting properties, dealing with the intersections of quadrics and flat spaces are obtained. Later, these results are applied to obtain filing schemes. In the later parts of the research, a mathematical (linear) model (representation) for multiple-valued attributes is developed. The design of this (representation model) depends on matrices, over a finite field, with a certain property. A method, applying the theory of spreads, for obtaining some of these matrices, is given. Finally, using this representation, a filing scheme for multiple-valued attributes is obtained.

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

It is a pleasure and privilage to acknowledge my sincerest gratitude to Professor R. C. Bose, not only for proposing the problems in the dissertation and making numerous suggestions and comments during its progress, but also for his excellent guidance and encouragement. It has been a pleasure to work under him.

I wish to express my special thanks to Professor T. A. Dowling for his encouragement and inspiration during my graduate program here. I also express my thanks to him for reading the manuscript very carefully and suggesting many valuable improvements.

To Professors R. R. Kuebler, Jr., D. G. Kelly, K.J.C. Smith and G. A. Magó, the other members of my doctoral committee I express my thanks, for their review of my work and suggestions for improving it. I would also like to thank Professor I. M. Chakravarti for agreeing to be a member of my committee.

I wish to express my thanks to Professor N. L. Johnson, Chairman and the Department of Statistics for providing the financial assistantship. To Professor Deal, Chairman, Department of Mathematics, Colorado State University, Fort Collins, I express my thanks for provinding the research facilities during my stay there.

To the members of the faculty and staff of the Department of Statistics who have contributed to my graduate training I wish to express my sincere gratitude.

I owe a deep sense of gratitude to my parents and brothers for their active cooperation and enthusiam. To my grandfather, Dr. N.S.R.

Sastry, I express my thanks for encouraging me into the field of Statistics.

For her excellent typing of the manuscript I extend my thanks to Mrs. Susan Drum.

# SUMMARY

The development of high-speed computers has produced an "information revolution" in its sources and as well as its types. Computers have provided means to comprehend with large volumes of data. But the question naturally arises as to how one may best file information in computer storage for further use. The efficiency of such a filing scheme can be evaluated in terms of the time it requires to satisfy a query. In this research we apply combinatorial methods to obtain efficient filing schemes.

Until recently, the best known type of filing scheme has been the inverted filing scheme. It is characterized by a one-one correspondence between the buckets and the levels of the attributes. These schemes allow efficient retrieval of queries of size one. But for higher order queries, it involves numerous cross comparisons.

We introduce the necessary terminology, along with the mathematical formulation of the problem, which was motivated by Ray-Chaudhuri [31][†], in Chapter I. Then a brief review of relevant combinatorial techniques is given. Finally, a brief review of the basic filing schemes with examples is given. Chapter II introduces the filing schemes that are derived from orthogonal arrays and partially balanced arrays. Also, some improved versions of existing filing schemes are given.

Chapter III deals mainly with quadrics and the filing schemes that are derived from them. A few basic results about the intersections of quadrics with (N-2)-flats are obtained. These results are later applied

---

[†] The number in square brackets refer to references listed in the bibliography at the end.

to obtain filing schemes for multiple-valued attributes.

Finally, in Chapter IV, we develop a linear representation for multiple-valued attributes. This representation enables us to develop filing schemes for multiple-valued attributes. These filing schemes depend on matrices with property $R_t(r_1, r_2, \ldots, r_\ell)$. For $t=2$, we give a construction procedure for optimal matrices. For higher values of $t$, it becomes a difficult problem.

# CHAPTER I

# INTRODUCTION

"It is unworthy of excellent men to lose hours like slaves in the labor of calculations which could safely be relegated to anyone else if machines were used."

Leibnitz 1671

With the advent and advancement of computer technology there has been a proliferation in information collection - its sources as well as its types. With the growing complexity and abundance of information we are faced not just with the problem of storing and retrieving a few words and facts but rather of storing and retrieving billions of words and facts. In an attempt to retrieve information from such a file we could go through these billions of facts - be they five or fifty billions - serially. But a simple computation will show that if a person spends his entire life from the day he turned twenty until the day he died, he still could not read one billion words. In light of this problem, the individual does not know where to begin to look. Our job, then, is to structure the files to assist him. The question naturally arises as to how one may best file information for further use. The success of any filing scheme can be measured in terms of the ease with which it is possible to retrieve information pertinent to a given task. With the present high speed computer technology the memory (storage) can be advantageously utilized to store well-organized files which can be retrieved rapidly. This fact provides a basis for the formulation of problems of designing filing schemes for efficient information retrieval. The actual construction of such files poses a variety of questions some of which can be tackled by combinatorial mathematics. This is the line of approach which we will follow in this research.

## 1.1 Terminology

For the purpose of this dissertation, a *file* is a collection of related records. A *record*, in turn, is comprised of two parts. The first part consists of an identifier number, which is unique with respect to a record. This identification is known as the *primary key*. Sometimes the records are also assigned a second identifier known as the *secondary key*. This is not unique. The second part consists of a number of related data fields. These data fields correspond to the values of a number of information variables, also known as *attributes*. We will assume that each attribute can take only one of finitely many different values. Hence the data field of a record consists of a particular combination of levels of attributes associated with the record. There may be more than one record associated with the same data field. Once the records are obtained, they can be stored in some permanent location (memory). The identifier of this location is called an *accession number*. One aspect of the basic problem of file organization is the definition of the correspondence between accession numbers and primary keys. This process is called *key transformation*. This has been discussed by a number of researchers in computer systems. For additional details and bibliography, the reader is referred to Buchholz [16]. We assume that the accession numbers have already been assigned by some procedure. From now on we deal with accession numbers alone, as they are smaller in size and easier to handle than the records.

An *address* is usually a number compounded of two or more coordinates that physically select the location. As an example, on a disk file, various digit groups of an address might specify position of a track, track, disk side, and module (group). A collection of addresses is called a *bucket*. The buckets form a partition of the addresses assigned for storage. The accession number of the records are stored at the addresses. We assume that in a bucket, exactly one accession number is stored at one address.

Finally, a *query* is a request to retrieve records containing certain information. A query is represented by a vector. The number of components of the vector indicate the size or order of the query.

## 1.2 Formulation of the problem

In this section, a mathematical model for filing schemes will be formulated. The approach used is similar to that of Ray-Chaudhuri [31] for the case in which retrieval pertains to only one level of $\ell$ attributes.

Let $A_{ij}$ denote the j-th level of the i-th attribute, where $j = 0,1,2,\ldots,n_i-1$; $i = 1,2,\ldots,\ell$. Also let

$$A_i = \{A_{ij}: \ j = 0,1,2,\ldots,n_i-1\} \qquad (1.2.1)$$

and

$$\Omega = \{A_{ij}: \ j = 0,1,\ldots,n_i-1; \ i = 1,2,\ldots,\ell\}. \qquad (1.2.2)$$

That is, $A_i$ is the set of all levels of the i-th attribute and $\Omega$ is the collection of all levels of all attributes.

A file $F$ is denoted by a triplet, $(\Lambda, \Omega, f)$ where $\Lambda$ is the collection of all records; $\Omega$ the set of all attribute levels; and $f$ a mapping from $\Lambda$ to $\Omega$ such that for any record $I \in \Lambda$,

$$f(I) = (A_{1j_1}, A_{2j_2}, \ldots, A_{\ell j_\ell}) \tag{1.2.3}$$

where $0 \leq j_k \leq n_k - 1$ for all $k$. That is to say $f(I)$ indicates the attribute levels possessed by the I-th record. Since each record contains exactly one level of each attribute we have

$$|f(I) \cap A_i| = 1, \qquad \text{for all } i \tag{1.2.4}$$

The storage rule $S$ of the file $F$ is a triplet $(\Lambda, M, \sigma)$ where $\Lambda$ is the collection of all records; $M$ the set of all possible integers corresponding to the set of possible addresses; $\sigma$ is a mapping from $\Lambda$ to subsets of $M$. The subset $\sigma(I)$ contains the addresses, where the accession number of the I-th record is stored. We implicitly assume that a single address is sufficient for storing any accession number. The number of addresses in $\sigma(I)$ indicates the number of repetitions of the accession number of the I-th record in the storage. Define

$$\zeta = \sum_{I \in \Lambda} |\sigma(I)| / |\Lambda| \; . \tag{1.2.5}$$

$\zeta$ is called the *redundancy* of the file. It depends on the storage rule.

The retrieval rule $R$ is a triplet $(\Delta, M, r)$ where $\Delta$ represents a class of subsets of $\Omega$; $M$ represents the set of addresses available for storage; and $r$ is a mapping from $\Delta$ to subsets of $M$ such that

$$|\sigma(I) \cap r(A)| = 1 \quad \text{if} \quad f(I) \supset A, \quad \text{where} \quad A \in \Delta. \tag{1.2.6}$$

In other words, only one of the addresses, where the accession number of I-th record is stored, is related to the retrieval of the query A. $\Delta$ is called the set of queries. The filing system is said to be of order $t$ if for each $A$ belonging to $\Delta$ the relation

$$|A| \leq t$$

holds. The time required to retrieve the records pertaining to a query $A$ is denoted by $\tau(A)$. The *average retrieval time* is

$$\tau = \sum_{A \in \Delta} \tau(A)/|\Delta| \quad . \tag{1.2.7}$$

$\tau$ depends on the retrieval rule.

A filing scheme is completely specified by indicating $F$, $S$ and $R$. The scheme will be called *optimal* if $\zeta$ and $\tau$ are minimum. But it is not always possible to obtain optimum filing schemes as $\zeta$ and $\tau$ are somewhat inversely proportional to each other. So the best way to tackle the problem will be by arbitrarily fixing one and minimizing the other. In this research, though we do not specify the values of $\zeta$ and $\tau$ we aimed at minimizing $\tau$ for reasonable values of $\zeta$.

As an example let us consider a simple filing scheme, known as the *inverted filing system.* This scheme is characterized by a one-one correspondence between buckets and levels of the attributes. Following the same notation as above, corresponding to each $A_{ij}$ ($j = 0,1,\ldots,n_i-1$; $i = 1,2,\ldots,\ell$) a bucket $M_{ij}$ (a set of addresses in the memory) is assigned. Let $W_{ij}$ be the *identification number* of $M_{ij}$. $W_{ij}$ can be taken as the first address in $M_{ij}$ if all the addresses are in a sequence or we can define the identifier $W_{ij}$ of $M_{ij}$ as

$$W_{ij} = \sum_{k=0}^{i-1} n_k + j, \qquad \text{where } n_0 = 0 \, . \tag{1.2.8}$$

A record $I$ is stored in the bucket $M_{ij}$ if the record contains the level $A_{ij}$. Since a record contains $\ell$ levels - one level of each attribute - it will be stored in $\ell$ buckets. That is

$$\sigma(I) \cap M_{ij} \neq \emptyset \quad \text{if} \quad f(I) \supset A_{ij} \tag{1.2.9}$$

and

$$|\sigma(I)| = \ell \ . \tag{1.2.10}$$

The retrieval rule for queries of size one is very simple. As an example, suppose we want to retrieve all records containing $A_{22}$. Then first, the identification number of the bucket corresponding to $A_{22}$ is calculated from (1.2.8),

$$W_{22} = n_1 + 2 \ .$$

By matching $W_{22}$ with the identification numbers of the buckets, the bucket $M_{22}$ is determined. $M_{22}$ gives all the accession numbers of records containing the level $A_{22}$. So, this retrieval rule involves a simple arithmatic calculation and a simple matching. The dominant factor, in terms of time is the matching of identification numbers. When the queries $A_{ij}$ are equally likely, we can show that the average retrieval time, $\tau$, using serial comparison is

$$\tau = \frac{(b+1)\tau_b}{2} \tag{1.2.11}$$

where $b = \sum_{i=1}^{\ell} n_i$ and $\tau_b$ represents the time required for each comparison. If we use binary search technique (for matching) instead of serial comparison, then

$$\tau = [\log_2 b]_+ \ \tau_b \tag{1.2.12}$$

where $[u]_+$ is the smallest integer greater than $u$.

Though the filing scheme appears simple for queries of size one, in terms of retrieval procedure, it becomes complicated when higher order queries are involved. For example the retrieval of records containing two levels $A_{11}$ and $A_{22}$ involves three steps - i) identify the buckets

$M_{11}$ and $M_{22}$ ii) extract all records contained in the buckets and iii) find the records common to both $M_{11}$ and $M_{22}$ by making cross comparisons. So the dominant factor in retrieval time will be the time required to make cross comparisons. This will depend on the sizes of the buckets. This fact represents the most striking disadvantage of the inverted filing system for retrieving records pertinent to two-fold queries. For higher order queries, this problem becomes progressively more and more serious.

The inverted filing system has been generalized to accommodate higher order queries. This scheme is also characterized by a one-one correspondence between queries and buckets. But it has enormously large redundancy. Further the retrieval time depends on the number of buckets. These considerations suggest the need to look for new filing schemes. In the remaining sections we shall consider the problem of construction of filing schemes by using various methods of combinatorial mathematics.

## 1.3 Finite Geometries

Abraham, Ghosh, and Ray-Chaudhuri [1] and Ghosh and Abraham [23] have applied combinatorial methods to the construction of some efficient filing schemes. To obtain these results, properties of finite geometries are extensively used. We also use these properties in the following chapters. So, we shall briefly summarize the properties of the two types of finite geometries: the finite projective geometry, denoted by $PG(N,q)$ and the finite Euclidean geometry, denoted by $EG(N,q)$ where $q$ is a prime power.

Projective geometry:

In a finite projective geometry $PG(N,q)$ of dimension $N$ based on a Galois field $GF(q)$, where $q$ is a prime power, the points can be taken as (N+1)-tuples $\underline{x} = (x_0, x_1, \ldots, x_N)$ where $x_0, x_1, \ldots, x_N$ are elements of GF(q). The (N+1)-tuple $\delta\, \underline{x} = (\delta x_0, \delta x_1, \ldots, \delta x_N)$ is regarded as the same point as $\underline{x}$ for any non-zero element $\delta$ of GF(q). The (N+1)-tuple (0,0,...,0) is not regarded as a point of PG(N,q). So a point P is uniquely determined by a non-null vector $\underline{x}$ and conversely a point P uniquely determines a vector $\underline{x}$ up to an arbitrary non-zero multiple $\delta$ of GF(q). An m dimensional flat space in PG(N,q) is defined by the set of points which satisfy (N-m) independent linear homogeneous equations

$$\begin{array}{c} a_{10}x_0 + a_{11}x_1 + \ldots + a_{1N}x_N = 0 \\ a_{20}x_0 + a_{21}x_1 + \ldots + a_{2N}x_N = 0 \\ \vdots \qquad\qquad \vdots \\ \vdots \qquad\qquad \vdots \\ a_{N-m0}x_0 + a_{N-m1}x_1 + \ldots + a_{N-mN}x_N = 0 \end{array} \tag{1.3.1}$$

or

$$\underline{A}\, \underline{x}^T = 0 \tag{1.3.2}$$

where $\underline{A}$ is a (N-m)×N matrix with the elements $a_{ij}$ from GF(q) and $\underline{x}^T$ is the transpose of $\underline{x}$. Thus the points which satisfy one linear homogeneous equation define an (N-1)-flat in PG(N,q). A point in PG(N,q) satisfies N independent linear homogeneous equations. Hence a point is called a 0-flat, a line a 1-flat, a plane a 2-flat and so on.

Let $\phi(N,m,q)$ denote the number of m-flats in PG(N,q), then

$$\phi(N,m,q) = \frac{(q^{N+1}-1)(q^N-1)\ \ldots\ (q^{N-m+1}-1)}{(q^{m+1}-1)(q^m-1)\ \ldots\ (q-1)} \tag{1.3.3}$$

The function $\phi$ satisfies the following property

$$\phi(N,m,q) = \phi(N,N-m-1,q) \qquad (1.3.4)$$
$$\phi(N,-1,q) = 1 \quad \text{(by convention)}.$$

If

$$\underset{(N-m)\times(N+1)}{\underline{A}} \quad \underset{(N+1)\times 1}{\underline{x}^T} = \underset{(N-m)\times 1}{\underline{0}}$$

represents an m-flat then all points whose corresponding row vectors lie in the vector space generated by the rows of $\underline{A}$ constitute an (N-m-1)-flat which is called the *dual* of the m-flat. Using the property of duality it can be shown that the number of m-flats containing (passing through) a given t-flat $(N \geq m \geq t)$ is equal to the number of (N-m-1)-flats contained in a given (N-t-1)-flat, i.e. $\phi(N-t-1,N-m-1,q)$.

Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ be $k$ (N+1)-tuples. They are said to be dependent if there exists elements $c_1, c_2, \ldots, c_k$ in GF(q), not all simultaneously zero, such that

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_k\underline{x}_k = \underline{0} \,. \qquad (1.3.5)$$

Otherwise they are said to be independent. A set of points of the projective geometry are said to be dependent or independent according as the corresponding row vectors are dependent or independent.

Let $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_m$ be a set of $(m+1)$ independent solutions of the equations in (1.3.1). Then any solution of (1.3.1) can be expressed as a linear combination of the vectors $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_m$. Hence any point of $\sum_m$ an m-dimensional flat, can be expressed as a linear combination of the points $P_0, P_1, \ldots, P_m$ corresponding to $\underline{x}_0, \underline{x}_1, \ldots, \underline{x}_m$. Finally by suitable linear transformation the row vectors of any (N+1) independent points in PG(N,q) can be taken as

$$\underline{x}_i = (\underbrace{0, \ldots, 0}_{i}, 1, \underbrace{0, \ldots, 0}_{N-i}) \quad , \quad i = 0,1,\ldots,N.$$

Let $\sum_m$ and $\sum_n$ be any two flats of dimension m and n respectively, in PG(N,q). The set of points common to both form a k-flat $\sum_k$, where $-1 \le k \le \min(m,n)$. That is

$$\textstyle\sum_k = \sum_m \cap \sum_n .$$

$\sum_k$ is called the *intersection* of $\sum_m$ and $\sum_n$. Let $\sum_h$ be the h-flat with lowest dimensions, containing both $\sum_m$ and $\sum_n$. $\sum_h$ is unique and is called the *join* of $\sum_m$ and $\sum_n$. Then

$$\max(m,n) \le h \le N$$

and

$$m + n = k + h .$$

Euclidean geometry:

A point in an N-dimensional finite Euclidean geometry EG(N,q) based on the Galois field GF(q) is defined to be an ordered N-tuple $(x_1, x_2, \ldots, x_N)$ where $x_i$ is an element of GF(q) for all i. The N-tuple (0, 0, ..., 0) is also a point of EG(N,q).

The m-flats $(0 \le m \le N-1)$ of EG(N,q) are defined by non-homogeneous equations. The set of points which satisfy (N-m) independent and consistent linear equations form an m-flat

$$\begin{array}{c} a_{10} + a_{11}x_1 + \ldots + a_{1N}x_N = 0 \\ a_{20} + a_{21}x_1 + \ldots + a_{2N}x_N = 0 \\ \vdots \qquad\qquad\qquad \vdots \\ a_{N-m,0}+a_{N-m,1}x_1+\ldots+ a_{N-m,N}x_N = 0 \end{array} \tag{1.3.6}$$

or

$$\underline{A}\ \underline{x}^T = \underline{a}_0 \tag{1.3.7}$$

where $\underline{A}$ is a $(N-m)\times(N+1)$ matrix with elements from $GF(q)$. The number of points in a m-flat is $q^m$.

The Euclidean geometry $EG(N,q)$ can be obtained from the projective geometry $PG(N,q)$ by deleting the so-called $(N-1)$-flat at infinity $x_0 = 0$ and all points and flats contained in it. Hence the number of m-flats in $EG(N,q)$ is equal to the number of m-flats in $PG(N,q)$ less the number of m-flats contained in the $(N-1)$-flat $x_0 = 0$, i.e.,

$$\phi(N,m,q) - \phi(N-1, m,q) = q^{N-m} \phi(N-1,m-1,q). \tag{1.3.8}$$

The various m-flats can be partitioned into parallel bundles by allowing the associated vectors $\underline{a}_0$ (in 1.3.7) to assume all possible values. In this way, there are $q^{N-m}$ m-flats in each such parallel bundle and $\phi(N-1, m-1, q)$ distinct parallel bundles in all. Finally each point in $EG(N,q)$ lies in exactly one of the m-flats belonging to any parallel bundle.

## 1.4 Balanced filing schemes

Abraham, Ghosh and Ray-Chaudhuri [1] were the first to use finite geometries to construct combinatorial filing schemes. Their method was developed for binary-valued attributes. It consisted of forming groups of recrods in such a manner that the group containing records pertaining to a given query could be determined algebraically, thus expediting the search. Ray-Chaudhuri [31] discussed some further combinatorial properties of file organization schemes for binary-valued attributes. Ghosh and Abraham [23] developed the theory for file organization schemes for multiple valued attributes, where the attributes have an equal number of possible values. However these schemes were limited to queries involving two values from two different attributes.

As an example we consider a filing scheme obtained by Abraham, Ghosh, and Ray-Chaudhuri [1]. $A_1, A_2, \ldots, A_\ell$ be $\ell$ attributes each with two levels. Consider a $PG(N,q)$ such that $\ell = \phi(N,0,q)$. Then the points of the geometry are identified with the attributes, the correspondence being one-one. We assume that the retrieval pertains to only one level of each attribute. The buckets are identified with the lines of the geometry. So the number of buckets is

$$b = \phi(N,1,q)$$

and the size of the bucket is $\phi(1,0,q) = q+1$. Since any two points will lie on exactly one line of the geometry, any pair of attributes will belong to exactly one bucket. In any bucket there are $\frac{q(q+1)}{2}$ pairs. So each pair corresponds to exactly $\binom{q}{2}$ queries.

Theorem (1.4.1). Given that the retrieval pertains to exactly one level of each attribute, there exists a filing scheme for queries of size two with $\ell = \phi(N,0,q)$ attributes each with two levels, (i.e. $n_1 = n_2 = \ldots = n_\ell = 2$) and with $b = \phi(N,1,q)$ blocks.

Example (1.4.1). Suppose there are $\ell = 7$ attributes each with two levels. Consider the projective geometry $PG(2,2)$. The points of this geometry are triplets. For convenience, we shall write them as $x_0x_1x_2$. The lines of the geometry are

$$x_0 = 0;\ x_1 = 0;\ x_2 = 0;\ x_0+x_1 = 0;\ x_0+x_2 = 0;$$
$$x_1+x_2 = 0;\ \text{and}\ x_0+x_1+x_2 = 0.$$

The attributes, and, together with the corresponding points of $PG(2,2)$ are

$$A_1 = 001,\ A_2 = 010,\ A_3 = 011,\ A_4 = 100$$
$$A_5 = 101,\ A_6 = 110,\ A_7 = 111\ .$$

The buckets are constructed by storing in them the accession numbers of records which have the following pairs of attributes.

| Bucket No. | Attribute pairs | Identification |
|---|---|---|
| 1 | $A_1A_2$; $A_1A_3$; $A_2A_3$ | 100 |
| 2 | $A_1A_4$; $A_1A_5$; $A_4A_5$ | 010 |
| 3 | $A_2A_4$; $A_2A_6$; $A_4A_6$ | 001 |
| 4 | $A_1A_6$; $A_1A_7$; $A_6A_7$ | 110 |
| 5 | $A_2A_5$; $A_2A_7$; $A_5A_7$ | 101 |
| 6 | $A_3A_4$; $A_3A_7$; $A_4A_7$ | 011 |
| 7 | $A_3A_5$; $A_3A_6$; $A_5A_6$ | 111 |

Each bucket has an identification number, which is taken to be the triplet of the coefficient of the equation of the line corresponding to the bucket. Within each bucket the accession numbers of records are subdivided into subgroups called subbuckets corresponding to each pair of attributes. The identification number of each subbucket is obtained by concatenation of the binary representation of the two attributes corresponding to it. Thus the subbuckets are represented as follows:

| Bucket Id. | Subbucket Id. | Accession numbers |
|---|---|---|
| 100 | 001 010 | |
| | 001 011 | |
| | 010 011 | |
| 010 | 001 100 | |
| | 001 101 | |
| | 100 101 | |

so on.

In storing the accession numbers of the records in the subbuckets of any bucket, it is likely that the same accession number is entitled to be stored in more than one subbucket, but this is avoided by storing it in the first subbucket it is entitled to. Suppose a record is represented by $(A_1, A_2, A_3, A_4, \overline{A}_5, \overline{A}_6, \overline{A}_7)$ (where $A_i$ indicates the presence of $A_i$; $\overline{A}_i$ indicates the absence of $A_i$) then the accession number of this record is stored in the subbucket 001 010 and not in 001 011 or 010 011 under the bucket 100. This accession number, however, is stored again in the subbucket 001 100 of the bucket 010; 010 100 of 001 and 011 100 of 011. Thus it is obvious that in this scheme any accession number is stored more than once, and this is the price that has to be paid for fast retrieval.

Suppose the following query was posed, "all records with $A_2$ and $A_4$ are to be retrieved". Then the search procedure is as follows: The attributes $A_2$ and $A_4$ are converted to points 010 and 100 respectively and the equation of the line containing these two points in a PG(2,2) is determined by substituting the coordinates of these two points in the general equation

$$a_0x_0 + a_1x_1 + a_2x_2 = 0$$

and solving for $a_0$, $a_1$ and $a_2$ in GF(2). On substitution, we have $a_0 = 0$ and $a_1 = 0$, hence the related line is $x_2 = 0$ and the identification number of the corresponding bucket is 001. Inside this bucket, the subbucket with identification number 010 100 is searched and this happens to be the first subbucket number in the bucket 001. Hence all the accession numbers in this subbucket are the required accession numbers. If the query was to find all the accession numbers of the records which have $A_2$ and $A_6$, then the search procedure would lead to the second subbucket, namely 010 110 in 001. In that case, all the accession numbers in this subbucket have to be retrieved and the accession numbers in the previous subbucket have to be searched to find $A_2 A_6$. This can be made an easy task by grouping before hand the accession numbers in the subbucket 010 100 into two groups, namely, the accession numbers of $A_2A_4A_6$, and the remaining and then using the *chaining* technique between subbucket 010 110 and the group of accession numbers under $A_2A_4A_6$ in the subbucket 010 100.

The retrieval time of the records pertaining to a query for the above filing scheme can be split up into the following components:

1. $T_1$ = time needed to solve the algebraic equation to determine the bucket,
2. $T_2$ = time needed for matching the bucket identification number,
3. $T_3$ = time needed for matching the subbucket identification number,
4. $T_4$ = time needed for chaining the subbuckets, if needed.

Hence the total retrieval time for any query A is

$$\tau(A) = T_1 + T_2 + T_3 + T_4 .$$

## 1.5 Combinatorial Configurations

A *combinatorial configuration* $(\Omega,k,\Delta,b)$ consists of a master set $\Omega$ (the set of all attribute levels), a class of subsets $\Delta$ (the set of queries) and blocks (buckets) $B_1, B_2, \ldots, B_b$ (which are certain subsets of $\Omega$) such that

i) $B_h \leq k$, for all $h$,

ii) for every $A \in \Delta$, there exists an $h$ such that $A$ is contained in $B_h$.

If $|A| \leq t$ for all $A$ in $\Delta$, then the configuration is said to be of order $t$, and is known as an $(\Omega,k,t,b)$ configuration. The actual construction of $(\Omega,k,t,b)$ configurations with minimum $b$ is a very difficult problem in combinatorial mathematics. For the case of $t = 2$, and $n_1 = n_2 = \dots = n_\ell = q$ such configurations are essentially equivalent to certain group divisible (GD) designs used in statistics. In most situations, optimal schemes are largely unknown and perhaps can be found through systematic trial and error.

A *combinatorial filing scheme* may be based on a combinatorial configuration. The blocks $B_1, B_2, \dots, B_b$ are arranged in serial order. For each $A$ in $\Delta$, define $\gamma(A) = h$ if $A$ is contained in $B_h$ but is not contained in $B_{h'}$, for $h' < h$. Hence $B_h$ is the first block which contains $A$. Let $\beta_h$ denote the collection of all subsets $A$ of $\Omega$ such that $\gamma(A) = h$. To each combination of $A$ and $h$ let there correspond sufficiently large disjoint subsets $M_{h,A}$ of $M$. The accession number of the I-th record is stored in an element of $M_{h,A}$ if and only if the largest set which $f(I)$ has in common with $B_h$ is the subset $A$ in $\beta_h$; i.e., if

$$f(I) \cap B_h = A.$$

Let

$$M_h = \bigcup_{A \in \beta_h} M_{h,A} .$$

The sets $M_h$ may be called the buckets of the filing scheme while the subsets $M_{h,A}$ may be called the subbuckets.

The retrieval procedure for any query simply involves the determination of the appropriate bucket by identifying the first block which contains the subset specified in the query. Then, all subbuckets corresponding to subsets which contain the query set are located and the accession numbers obtained. Thus, the retrieval function may be formally written as

$$r(A) = \bigcup_{A \subseteq C \in \beta_h} M_{h,C}$$

where $A \in \Delta$, and $\gamma(A) = h$. From the preceeding remarks, one can see that once a combinatorial configuration has been constructed, a reasonable filing scheme may be readily based on it. In particular, Bose, Abraham and Ghosh [7] have used a procedure similar to this. Finally the concepts of combinatorial configuration and combinatorial filing schemes that are described here are equivalent to the ones considered by Ray-Chaudhuri [31] for the situation in which only one level of any attribute was of interest with respect to retrieval.

## 1.6 Second order combinatorial configurations

The problem of constructing second order combinatorial configurations is essentially the same as that of constructing certain incomplete block designs used in statistical research. Of special interest are balanced incomplete block designs and group divisible designs. The combinatorial properties of these designs have been studied by a number of researchers. For further reference we refer to Bose [3], Bose [4], Bose, Shrikhande and Bhattacharya [12], Rao [28], and Sprott [33].

A *balanced incomplete block* (BIB) design with parameters $(v,b,r,k,\lambda)$, is an arrangement of $v$ objects into $b$ subsets, called blocks, such that

i) each block contains exactly $k$ objects,

ii) each object occurs in $r$ distinct blocks,

and iii) each pair of objects occurs together in $\lambda$ distinct blocks.

Assuming that the retrieval pertains to only one level of an attribute the problem of construction of a combinatorial configuration

$(\Omega,k,2,b)$ where $\Omega = \{A_1,A_2,\ldots,A_\ell\}$ is exactly the same as the construction of a BIB design with parameters $(\ell,b,r,k,\lambda=1)$. For any given value of $k$ the BIB design, if it exists, has the minimum number of blocks. Hence the corresponding combinatorial configuration is optimal. There may not exist BIB designs for all possible choices of $v$, $b$, $r$, $k$ and $\lambda$. Some existing designs are, for $\lambda = 1$.

| v | b | r | k |
|---|---|---|---|
| 7 | 7 | 3 | 3 |
| 13 | 13 | 4 | 4 |
| 13 | 26 | 6 | 3 |
| 15 | 35 | 7 | 3 |
| 16 | 20 | 5 | 4 |
| 19 | 57 | 9 | 3 |
| 25 | 50 | 8 | 4 |
| 25 | 100 | 12 | 3 |
| 27 | 117 | 13 | 3 |
| 28 | 63 | 9 | 4 |
| 40 | 130 | 13 | 4 |
| 66 | 143 | 13 | 6 |
| 91 | 195 | 15 | 7 |
| 113 | 226 | 16 | 8 |
| 145 | 232 | 16 | 10 |
| 145 | 290 | 18 | 9 |

The above list is by no means complete. We shall delay the description of GD designs until the next chapter.

Example (1.6.1). Suppose there are seven attributes each with two levels. We shall assume that the retrieval pertains to only one level of each attribute. Let $A_1$, $A_2$, $A_3$, $A_4$, $A_5$, $A_6$ and $A_7$ be the attributes. The buckets are identified with the blocks of a BIB design with parameters (7,7,3,3,1). So the buckets are

$$B_1 = \{A_1, A_2, A_4\},\quad B_2 = \{A_2, A_3, A_5\},\quad B_3 = \{A_3, A_4, A_6\},$$
$$B_4 = \{A_4, A_5, A_7\},\quad B_5 = \{A_5, A_6, A_1\},\quad B_6 = \{A_6, A_7, A_2\},$$
$$B_7 = \{A_7, A_1, A_3\}\ .$$

The subbuckets are formed by considering all subsets in a bucket. That is,

$M_1$: $\{A_1\}$; $\{A_2\}$; $\{A_4\}$; $\{A_1A_2\}$; $\{A_1A_4\}$; $\{A_2A_4\}$; $\{A_1A_2A_4\}$,
$M_2$: $\{A_3\}$; $\{A_5\}$; $\{A_2A_3\}$; $\{A_2A_5\}$; $\{A_3A_5\}$; $\{A_2A_3A_5\}$,
$M_3$: $\{A_6\}$; $\{A_3A_4\}$; $\{A_3A_6\}$; $\{A_4A_6\}$; $\{A_3A_4A_6\}$,
$M_4$: $\{A_7\}$; $\{A_4A_5\}$; $\{A_4A_7\}$; $\{A_5A_7\}$; $\{A_4A_5A_7\}$,
$M_5$: $\{A_1A_5\}$; $\{A_1A_6\}$; $\{A_5A_6\}$; $\{A_1A_5A_6\}$,
$M_6$: $\{A_2A_6\}$; $\{A_2A_7\}$; $\{A_6A_7\}$; $\{A_2A_6A_7\}$,
$M_7$: $\{A_1A_3\}$; $\{A_1A_7\}$; $\{A_3A_7\}$; $\{A_1A_3A_7\}$.

The subset $A_2$ in $B_2$ is not considered as a subbucket in $M_2$, for

$$\gamma(A_2) = 1\ .$$

That is, a subset $A$ of $B_i$ is taken as a subbucket in the corresponding $M_i$ only if $\gamma(A) = i$.

Let $I$ be any record such that

$$f(I) = (A_1, \bar{A}_2, A_3, \bar{A}_4, A_5, \bar{A}_6, \bar{A}_7)$$

where $A_i$ indicates the presence of the i-th attribute and $\bar{A}_i$ indicates the absence of it. Then

$$f(I) \cap B_1 = \{A_1\}\ ,\quad \gamma(A_1) = 1\ ;$$
$$f(I) \cap B_2 = \{A_3A_5\},\quad \gamma(A_3A_5) = 2;$$
$$f(I) \cap B_3 = \{A_3\}\ ,\quad \gamma(A_3) = 2\ ;$$
$$f(I) \cap B_4 = \{A_5\}\ ,\quad \gamma(A_5) = 2\ ;$$
$$f(I) \cap B_5 = \{A_1A_5\},\quad \gamma(A_1A_5) = 5;$$
$$f(I) \cap B_6 = \{\emptyset\}\ ;$$

and

$$f(I) \cap B_7 = \{A_1A_3\}, \quad \gamma(A_1A_3) = 7.$$

Hence the I-th record accession number will be stored in the subbucket $\{A_1\}$ of the 1st bucket, in the subbucket $\{A_3A_5\}$ of the second bucket, in the subbucket $\{A_1A_5\}$ of the fifth bucket and in the subbucket $\{A_1A_3\}$ of the seventh bucket. That is a record is stored in the subbucket representing the maximal subset common to the record and the bucket. So when retrieving a query we may have to search in more than one subbucket. This can be taken care by chaining the appropriate subbuckets. For example in the first bucket, the subbucket $\{A_1\}$ will be chained to $\{A_1A_2\}$, $\{A_1A_4\}$ and $\{A_1A_2A_4\}$.

To retrieve the query $\{A_3\}$, first we have to determine the bucket that contains the set $A_3$. Since $\gamma(A_3) = 2$, the bucket under consideration is the second bucket. The subbuckets $\{A_3\}$, $\{A_2A_3\}$, $\{A_3A_5\}$ and $\{A_2A_3A_5\}$ contain all the accession numbers that satisfy the query. Similarly we can retrieve all records that satisfy queries of size one and two.

Another type of combinatorial configuration is obtained by Ray-Chaudhuri [31] using covers in finite projective spaces.

An m-flat $\sum_m$ in $PG(N,q)$ is said to cover a $(t-1)$-flat $\sum_{t-1}$ if $\sum_m \supseteq \sum_{t-1}$ where $N \geq m \geq t-1$. A class of m-flats $(\pi_1, \pi_2, \ldots, \pi_b)$ is defined to be a $(b,t,m)$ cover if every $(t-1)$-flat in $PG(N,q)$ is contained in at least one of the m-flats $\pi_h$ belonging to the class. A $(b,t,m)$ cover is called a minimum $(b,t,m)$ cover if it contains the minimum number $b_0 = b_0(N,t,m,q)$ of m-flats required to cover every $(t-1)$-flat.

An important result in this connection is obtained by Ray-Chaudhuri [31].

<u>Theorem (1.6.1)</u>. There exists a $(\Omega,k,t,b)$ combinatorial configuration for the case

$$\ell = (q^{N+1}-1)/(q-1), \quad k = (q^{m+1}-1)/(q-1),$$

$$b = b_0(N,t,m,q), \quad \text{where} \quad N \geq m \geq t-1,$$

and $q$ is a prime power. The retrieval pertains to only one level of each attribute.

Some additional methods for forming covers are obtained by Koch [25].

# CHAPTER II

# MISCELLANEOUS FILING SCHEMES

In this chapter we obtain miscellaneous combinatorial configurations. Some of these configurations are improvements over the existing ones. We give a simple construction procedure for these designs, beginning with a description of the existing ones.

## 2.1 Orthogonal arrays and Partially balanced arrays

The problem of constructing combinatorial configurations with $k = \ell$ is equivalent to the problem of forming an array of ordered $\ell$-tuples (in which each coordinate corresponds to a unique attribute) in such a way that every possible ordered combination of $t$ coordinates occurs at least once. For the case in which $\Omega$ consists of $\ell$ attributes, each with $n(n_1=n_2=\ldots=n_\ell=n)$ levels, and in which all t-tuples occur exactly once, such a construction is called an *orthogonal array* of strength $t$, constraints $\ell$, and index unity. It is represented by $(b,\ell,n,t)$. Such arrays have been studied by Bose and Bush [9], Bush [17] and many others. For large values of $\ell$, it becomes very difficult to construct orthogonal arrays with index unity.

But for our purpose we do not need every t-tuple be covered exactly once but at least once. This leads to the concept of partially balanced array, as defined by Chakravarti [18], [19].

**Definition (2.1.1)** A *partially balanced array* of strength $t$ in $b$ blocks, $\ell$ attributes with $n$ levels each, is equivalent to a $(b\times\ell)$ matrix in which among the rows of each $t$ column submatrix, every possible permutation of the values in the vector $(u_1,u_2,\ldots,u_t)$ occurs exactly $\lambda(u_1,u_2,\ldots,u_t)$-times, where the value of $\lambda(u_1,u_2,\ldots,u_t)$ does not depend upon which $t$ columns are chosen.

For our purpose we need those partially balanced arrays for which a majority of $\lambda(u_1,u_2,\ldots,u_t)$ are unity. As with orthogonal arrays, the problem of constructing partially balanced arrays for large $\ell$ with $\lambda$'s near unity is another very difficult problem.

## 2.2 Composition

One way of tackling the problem for higher values of $\ell$ is to use the method, composition as suggested by Koch [25]. This procedure involves two steps - i) construct efficient partially balanced arrays for small values of $\ell$ and then ii) expand these arrays to obtain higher values of $\ell$. The arrays so obtained for large $\ell$ may not be efficient. That is we may have more blocks than are necessary. Koch [25] applied this procedure to obtain configurations of order two, three, and four when the attributes had $q$ (a prime power) levels.

## 2.3 Configurations of order two with $k = \ell$

Let us assume that there are $\ell$ attributes each with two levels, i.e. $n = 2$. A combinatorial configuration of order two, with $k = \ell$, can be represented by a $(b\times\ell)$ matrix in which among the rows of each two column submatrix, each of the four possible ordered 2-tuples (00), (01), (10), (11) occurs at least once. In this section, a method for constructing such matrices will be studied.

Consider the following $(4\times3)$ matrix

$$\begin{matrix}111\\100\\010\\001\end{matrix} \qquad (2.3.1)$$

This represents an efficient configuration of order two for three attributes. Further $\lambda(00) = \lambda(01) = \lambda(10) = \lambda(11) = 1$. Similarly for four attributes an efficient configuration of order two is given by

$$\begin{matrix} 1111 \\ 1000 \\ 0100 \\ 0010 \\ 0001 \end{matrix} \tag{2.3.2}$$

It has $b = 5$; $\lambda(11) = \lambda(10) = \lambda(01) = 1$ and $\lambda(00) = 2$. In fact the above design is a partially balanced array of strength two. The matrices (2.3.1) and (2.3.2) have a systematic structure - namely, the first row consists of ones and every column has exactly $[\frac{b}{2}]$[†] ones. Generalizing this idea, we have

<u>Theorem (2.3.1)</u>. Given $\ell$ attributes each with two levels, there exists a configuration of order two with $b$ blocks of size $\ell$. Further

$$b \le \alpha,$$

where $\alpha$ is the least integer such that

$$\binom{\alpha - 1}{[\frac{\alpha}{2}] - 1} \ge \ell . \tag{2.3.3}$$

<u>Proof</u>. Let $G$ be a $(\alpha\times\ell)$ matrix such that

i) the first element in each column is "one",

ii) every column has exactly $[\frac{\alpha}{2}] - 1$ "ones",

and iii) all columns are distinct.

Let $\underline{C}_1$ and $\underline{C}_2$ be any two columns of $G$,

---

[†] $b$ is the number of rows and $[\frac{b}{2}]$ denotes the integral part of $\frac{b}{2}$.

$$\underline{C}_1 = \begin{bmatrix} 1 \\ c_{11} \\ c_{21} \\ \vdots \\ \vdots \\ \\ c_{\alpha-1,1} \end{bmatrix} \qquad \underline{C}_2 = \begin{bmatrix} 1 \\ c_{12} \\ c_{22} \\ \vdots \\ \vdots \\ \\ c_{\alpha-1,2} \end{bmatrix} \qquad \underset{\alpha\times 2}{\underline{C}} = \begin{bmatrix} 1 & 1 \\ c_{11} & c_{12} \\ c_{21} & c_{22} \\ \vdots & \vdots \\ \vdots & \vdots \\ \\ c_{\alpha-1,1} & c_{\alpha-1,2} \end{bmatrix} \qquad (2.3.4)$$

where $C_{ij} = 0$ or $1$ for all $i$ and $j$. Clearly (11) occurs at least once in $\underline{C}$. The ones in $\underline{C}_1$ and $\underline{C}_2$ do not coincide as $\underline{C}_1$ and $\underline{C}_2$ have exactly the same number of "ones" and are distinct. Hence (01) and (10) will each occur at least once in $\underline{C}$. Finally (00) occurs at least once as

$$2[\tfrac{\alpha}{2}] - 2 \leq \alpha - 2.$$

Hence G is a partially balanced array of strength two.

Example (2.3.1). Let $\ell = 9$.

If $\alpha = 6$ then $\binom{6-1}{3-1} = 10 \geq 9$.

The $(6\times 9)$ matrix is,

$$\begin{array}{l} 111\ 111\ 111 \\ 111\ 100\ 000 \\ 100\ 011\ 100 \\ 010\ 010\ 011 \\ 001\ 001\ 010 \\ 000\ 100\ 101 \end{array}$$

Example (2.3.2). Let $\ell = 12$.

Then $\binom{7-1}{3-1} = 15 > 12$.

So the design is

$$\begin{array}{cccc}
111 & 111 & 111 & 111 \\
111 & 110 & 000 & 000 \\
100 & 001 & 111 & 000 \\
010 & 001 & 000 & 111 \\
001 & 000 & 100 & 100 \\
000 & 100 & 010 & 010 \\
000 & 010 & 001 & 001
\end{array}$$

The columns are in lexicographic order.

In particular, the following table indicates the relative sizes of $b$ and $\ell$ .

| $\ell$ | b | $\ell$ | b | $\ell$ | b |
|---|---|---|---|---|---|
| 3 | 4 | 12 | 7 | 54 | 9 |
| 4 | 5 | 18 | 8 | 81 | 10 |
| 6 | 6 | 27 | 8 | 108 | 10 |
| 9 | 6 | 36 | 9 | 162 | 11 |

In the special case, when $\ell = 3^u$ we have $\alpha = 2u + 2$, for

$$\binom{2u + 1}{u} \geq 3\binom{2u - 1}{u - 1} .$$

Where as for the same case the composition procedure gives

$$b = 3u + 1.$$

## 2.4 Configurations of order two, with $k \neq \ell$

Construction of configurations of order two with $k \neq \ell$, is equivalent to the construction of certain group divisible designs.

**Definition (2.4.1).** A *group divisible design* is an arrangement of $V = \ell n$ objects, belonging to $\ell$ groups of $n$ objects each, into $b$ blocks such that

i) each block contains $k$ objects,

ii) each object occurs in $r$ distinct blocks,

iii) each pair of objects, belonging to the same group, occurs together in $\lambda_1$ blocks,

and iv) each pair of objects, belonging to different groups occurs together in $\lambda_2$ blocks.

So, a GD design with $\lambda_1 = 0, \lambda_2 = 1$ represents a combinatorial configuration of order two appropriate to the multi-level attribute case with $\ell$ attributes, each with n levels. These configurations are optimal in the sense that each pair of levels of different attributes is covered exactly once. Hence for a given value of k, b is minimum. But GD designs may not always exist. One simple method of construction of these designs is given by Bose, Shrikhande and Bhattacharya [12].

Theorem (2.4.1). By omitting a particular treatment $\theta$ from a BIB design with parameters $v^*, b^*, r^*, k^*$, and $\lambda^* = 1$ together with all the blocks containing $\theta$, we get a GD design with parameters

$$v = v^* - 1, \quad b = b^* - r^*, \quad r = r^* - 1, \quad k = k^*,$$
$$\ell = r^*, \quad n = k^* - 1, \quad \lambda_1 = 0, \quad \lambda_2 = 1.$$

In particular if we start with a BIB design belonging to the series OS1, with parameters $v = q^2$, $b = q^2+q$, $r = q+1$, $k = q$, $\lambda = 1$ (q is a prime power) we get a series of GD designs with

$$v = b = q^2-1, \quad r = k = q, \quad \ell = q+1, \quad n = q-1$$
$$\lambda_1 = 0, \quad \lambda_2 = 1 .$$

As an example, take $q = 3$. Then

$$v = b = 8, \quad r = k = 3, \quad \ell = 4, \quad n = 2, \quad \lambda_1 = 0, \quad \lambda_2 = 1 .$$

$2_0 \quad 3_1 \quad 4_1$

$3_0 \quad 4_1 \quad 1_0$

$4_0 \quad 1_0 \quad 2_0$

$$\begin{array}{ccc} 1_1 & 2_0 & 3_0 \\ 2_1 & 3_0 & 4_0 \\ 3_1 & 4_0 & 1_1 \\ 4_1 & 1_1 & 2_1 \\ 1_0 & 2_1 & 3_1 \end{array}$$

In the rest of this section we develop a method for constructing configurations of order two, when $k \neq \ell$.

Lemma (2.4.1). Given an integer $m$, there exists an $(N \times m)$ zero-one matrix such that in any two columns taken together the pairs (10), (01) occur at least once. Further

$$N \le \underset{\alpha}{\text{minimum}} \; \left\{ \alpha : \binom{\alpha}{[\frac{\alpha}{2}]} \ge m \right\}.$$

Proof. Choose any $m$ of the possible $\binom{\alpha}{[\frac{\alpha}{2}]}$ "ones". Since every column has exactly the same number of "ones" and are all distinct, the pairs (10), (01) will occur at least once in any two columns taken together.

Example (2.4.1)

$$\begin{array}{ccc} 1_0 & 2_0 & 3_0 \\ 1_0 & 2_1 & 4_1 \\ 1_1 & 3_0 & 4_1 \\ 2_1 & 3_0 & 4_0 \\ 1_1 & 2_1 & 3_1 \\ 1_1 & 2_0 & 4_0 \\ 1_0 & 3_1 & 4_0 \\ 2_0 & 3_1 & 4_1 \end{array}$$

This represents a configuration of four treatments each with two levels in blocks of size three.

Lemma (2.4.2). Given 4m treatments each with two levels there exists a configuration of order two with b blocks of size 3m, where

$$b \le \alpha + [\tfrac{\alpha}{3}]^{+} + 8 \cdot \qquad (2.4.1)$$

Proof. Define

$$W_i = [(i-1)m+1,\ (i-1)m+2,\ \ldots,\ [\ m] \qquad i = 1,2,3,4.$$

That is $W_1$ is an m-tuple consisting of the first m treatments and so on. Denote by $W_i^j$ the j-th level of all the m treatments in $W_i$. Then the first eight blocks can be taken as

$$\begin{array}{ccc}
W_1^0 & W_2^0 & W_3^0 \\
W_1^0 & W_2^1 & W_4^1 \\
W_1^1 & W_3^0 & W_4^1 \\
W_2^1 & W_3^0 & W_4^0 \\
W_1^1 & W_2^1 & W_3^1 \\
W_1^1 & W_2^0 & W_4^0 \\
W_1^0 & W_3^1 & W_4^0 \\
W_2^0 & W_3^1 & W_4^1
\end{array}$$

These eight blocks do not cover the pairs (01), (10) corresponding to the treatments within a $W_i$ $(i = 1,2,3,4)$. We need at most $\alpha$ blocks to cover these pairs (within a $W_i$), where

$$\binom{\alpha}{[\frac{\alpha}{2}]} \ge m \cdot \qquad \text{(from lemma 2.4.1)}$$

Hence

$$b \le 8 + \alpha + [\tfrac{\alpha}{3}]^{+}$$

where $[\frac{\alpha}{3}]^{+}$ (the least integer greater than or equal to $\frac{\alpha}{3}$) is the number of extra blocks needed as the block size is $3m$.

Example (2.4.2)

$$\ell = 12, \quad k = 9 \quad \text{and} \quad m = 3$$

Then $\alpha = 3, \quad \text{so} \quad b \le 8 + 3 + 1 = 12$

$$\begin{array}{ccc}
1_0\ 2_0\ 3_0 & 4_0\ 5_0\ 6_0 & 7_0\ 8_0\ 9_0 \\
1_0\ 2_0\ 3_0 & 4_1\ 5_1\ 6_1 & 10_1 11_1 12_1 \\
1_1\ 2_1\ 3_1 & 7_0\ 8_0\ 9_0 & 10_1 11_1 12_1 \\
4_1\ 5_1\ 6_1 & 7_0\ 8_0\ 9_0 & 10_0 11_0 12_0 \\
1_1\ 2_1\ 3_1 & 4_1\ 5_1\ 6_1 & 7_1\ 8_1\ 9_1 \\
1_1\ 2_1\ 3_1 & 4_0\ 5_0\ 6_0 & 10_0 11_0 12_0 \\
1_0\ 2_0\ 3_0 & 7_1\ 8_1\ 9_1 & 10_0 11_0 12_0 \\
4_0\ 5_0\ 6_0 & 7_1\ 8_1\ 9_1 & 10_1 11_1 12_1 \\
1_1\ 2_0\ 3_0 & 4_1\ 5_0\ 6_0 & 7_1\ 8_0\ 9_0 \\
1_0\ 2_1\ 3_0 & 4_0\ 5_1\ 6_1 & 10_1 11_0 12_0 \\
1_0\ 2_1\ 3_1 & 7_0\ 8_1\ 9_0 & 10_0 11_1 12_0 \\
4_0\ 5_0\ 6_1 & 7_0\ 8_0\ 9_1 & 10_0 11_0 12_1
\end{array}$$

Theorem (2.4.3). Given $4m$ attributes each with two levels there exists a combinatorial configuration of order two with $b$ blocks of size $3m$. If the configuration for $v = 2m$, $k = \frac{3m}{2}$ exists then

$$b \le \min\{b^{*}+3,\ \alpha + [\tfrac{\alpha}{3}]^{+} + 8\},$$

otherwise

$$b \le \alpha + [\tfrac{\alpha}{3}]^{+} + 8,$$

where $b^*$ is the number of blocks required for $v = 2m$ and $k = \frac{3m}{2}$ configuration of order two and $\alpha$ is an integer such that

$$\binom{\alpha}{[\frac{\alpha}{2}]} \geq m .$$

Proof. Suppose that the configuration C of order two with $v = 2m$, $k = \frac{3m}{2}$ and $b^*$ blocks exists. Then m is clearly even. We will extend this configuration C to a new configuration C* as follows. Replace every attribute C by a pair of attributes of C*; i.e. if $i_j$ denotes the j-th level of the i-th attribute in C then it is replaced by the pair $(i_j, 2m + i_j)$. The blocks so obtained will cover all 2-tuples except

$$\{(i_0, 2m+i_1), (i_1, 2m+i_0), i = 1,2,\ldots,2m\}.$$

These 2-tuples can be covered by the following three blocks.

$$\{1_0, 2_0, \ldots, \tfrac{3m}{2}{}_0, 3m+1_1, 2m+2_1, \ldots, \tfrac{7m}{2}, \}$$
$$\{1_1, 2_1, \ldots, m_1, \tfrac{3m}{2}+1_0, \tfrac{3m}{2}+2_0, \ldots, 2m_0, 2m+1_0, \ldots, 3m_0, \tfrac{7m}{2}+1_1, \ldots, 4m_1\}$$
$$\{\tfrac{m}{2}+1_1, \ldots, 2m_1, \tfrac{5m}{2} + 1_0, \tfrac{5m}{2} + 2_0, \ldots, 4m_0\}$$

Hence

$$b \leq b^* + 3.$$

But from lemma (2.4.2), we have

$$b \leq \alpha + [\tfrac{\alpha}{3}]^+ + 8,$$

where

$$\binom{\alpha}{[\frac{\alpha}{2}]} \geq m.$$

Hence the result.

Example (2.4.3). $v = 8$, $k = 6$ .

From example (2.4.1), we have the configuration of order two with parameters $v = 4$, $k = 3$, $b = 8$. So using this configuration we have

$$\begin{array}{ccc}
1_0\ 5_0 & 2_0\ 6_0 & 3_0\ 7_0 \\
1_0\ 5_0 & 2_1\ 6_1 & 4_1\ 8_1 \\
1_1\ 5_1 & 3_0\ 7_0 & 4_1\ 8_1 \\
2_1\ 6_1 & 3_0\ 7_0 & 4_0\ 8_0 \\
1_1\ 5_1 & 2_1\ 6_1 & 3_1\ 7_1 \\
1_1\ 5_1 & 2_0\ 6_0 & 4_0\ 8_0 \\
1_0\ 5_0 & 3_1\ 7_1 & 4_0\ 8_0 \\
2_0\ 6_0 & 3_1\ 7_1 & 4_1\ 8_1 \\
\hline
1_0\ 2_0 & 3_0\ 5_1 & 6_1\ 7_1 \\
1_1\ 2_1 & 4_0\ 5_0 & 6_0\ 8_1 \\
2_1\ 3_1 & 4_1\ 6_0 & 7_0\ 8_0
\end{array}$$

Hence

$$b \leq 11.$$

## 2.5 Filing schemes based on finite geometries

Consider a projective geometry $PG(N,2)$ of dimension $N$. Any $(N+1)$ points in $PG(N,2)$ are either dependent or independent. If a set of $(N+1)$ points is dependent then there exists at least one $(N-1)$-flat in $PG(N,2)$ such that it contains all the $(N+1)$ points. If they are independent, then we shall show that there exists an $(N-1)$-flat in $PG(N,2)$ such that its complement contains these $(N+1)$ independent points.

Let $P_0, P_1, P_2, \ldots, P_N$ be a set of $(N+1)$ independent points in $PG(N,2)$. Let $\underline{P}_i$ denote the row vector of $P_i$, $i = 0,1,\ldots,N$. We can find a non-singular linear homogeneous transformation such that the row vector of $P_i$ is

$$\underline{P}_i = (\underbrace{0,0,\ldots,0}_{i},1,\underbrace{0,\ldots,0}_{N-i}) \quad i = 1,2,\ldots,N. \qquad (2.5.1)$$

Consider the (N-1)-flat,

$$x_0 + x_1 + \ldots + x_N = 0 . \qquad (2.5.2)$$

None of the points $P_i$ satisfy the equation (2.5.2). Hence they belong to the complement with respect to PG(N,2) of the (N-1)-flat

$$x_0 + x_1 + \ldots + x_N = 0.$$

So,

Theorem (2.5.1). Any (N+1) points in PG(N,2) are either contained in a (N-1)-flat or contained in the complement of an (N-1)-flat.

Theorem (2.5.2). Given that retrieval pertains to only one level of any attribute, there exists a filing scheme with $(2^{N+1}-1)$ attributes, which is oriented toward (N+1)-tuple queries. It has $2^{N+2}-2$ buckets.

Proof. Identify the $2^{N+1}-1$ attributes with the points of the geometry PG(N,2). The buckets are identified with (N-1)-flats and their complements with respect to PG(N,2). The number of buckets is

$$\begin{aligned} b &= 2 \times \phi(N, N-1, 2), \\ &= 2^{N+2}-2. \end{aligned}$$

The size of a bucket is either $2^N-1$ or $2^N$. Clearly, this filing scheme covers all (N+1)-tuples.

A bucket is subdivided into subbuckets by forming all possible (N+1)-tuples. Some of these subbuckets may be duplicated. Hence in any bucket there are at most $\binom{2^N}{N+1}$ subbuckets. A subbucket occurs exactly once in the filing scheme if the corresponding (N+1)-tuple has at least N independent points.

Example (2.5.1). Consider PG(2,2). Let

$$A_1 = 001\,, \quad A_4 = 100\,,$$
$$A_2 = 010\,, \quad A_5 = 101\,,$$
$$A_3 = 011\,, \quad A_6 = 110\,,$$
$$A_7 = 111\,.$$

The buckets and their corresponding attributes are

| Bucket No. | Identification | Attributes | Subbuckets |
|---|---|---|---|
| B1 | 1000 | $A_1,A_2,A_3$ | $A_1A_2A_3$ |
| B2 | 1001 | $A_4,A_5,A_6,A_7$ | $A_4A_5A_6$<br>$A_4A_5A_7$<br>$A_4A_6A_7$<br>$A_5A_6A_7$ |
| B3 | 0100 | $A_1,A_4,A_5$ | $A_1A_4A_5$ |
| B4 | 0101 | $A_2,A_3,A_6,A_7$ | $A_2A_3A_6$<br>$A_2A_3A_7$<br>$A_2A_6A_7$<br>$A_3A_6A_7$ |
| $B_5$ | 0010 | $A_2,A_4,A_6$ | $A_2A_4A_6$ |
| $B_6$ | 0011 | $A_1,A_3,A_5,A_7$ | $A_1A_3A_5$<br>$A_1A_3A_7$<br>$A_1A_5A_7$<br>$A_3A_5A_7$ |

| Bucket No. | Identification | Attributes | Subbuckets |
|---|---|---|---|
| B7 | 1100 | $A_1, A_6, A_7$ | $A_1A_6A_7$ |
| B8 | 1101 | $A_2, A_3, A_4, A_5$ | $A_2A_3A_4$<br>$A_2A_3A_5$<br>$A_2A_4A_5$<br>$A_3A_4A_5$ |
| B9 | 1010 | $A_2, A_5, A_7$ | $A_2A_5A_7$ |
| B10 | 1011 | $A_1, A_3, A_4, A_6$ | $A_1A_3A_4$<br>$A_1A_3A_6$<br>$A_1A_4A_6$<br>$A_3A_4A_6$ |
| B11 | 0110 | $A_3, A_4, A_7$ | $A_3A_4A_7$ |
| B12 | 0111 | $A_1, A_2, A_5, A_6$ | $A_1A_2A_5$<br>$A_1A_2A_6$<br>$A_1A_5A_6$<br>$A_2A_5A_6$ |
| B13 | 1110 | $A_3, A_5, A_6$ | $A_3A_5A_6$ |
| B14 | 1111 | $A_1, A_2, A_4, A_7$ | $A_1A_2A_4$<br>$A_1A_2A_7$<br>$A_1A_4A_7$<br>$A_2A_4A_7$ |

Notice that none of the subbuckets are duplicated. This is because any two points are independent and we are considering only three points. Suppose we are interested in retrieving records with attributes $\{A_1A_2A_4\}$. The corresponding points of the geometry are

$$A_1 = 001, \quad A_2 = 010, \quad A_4 = 100.$$

The equation of the complement of any 1-flat is

$$a_0x_0 + a_1x_1 + a_2x_2 = 1,$$

as we are dealing with PG(N,2). Hence the equation of any bucket is of the form

$$a_0x_0 + a_1x_1 + a_2x_2 = a_3 .$$

So the bucket containing (001), (010), (100) is obtained by solving the following equations

$$a_0 = a_3$$
$$a_1 = a_3$$
$$a_2 = a_3$$

Hence $a_3 = 1$. So the bucket is 1111. Similarly we can find the bucket corresponding to any query of size three.

## 2.6 Multiple level attributes

The filing schemes described in the previous sections deal mainly with the case $n_1 = n_2 = \cdots = n_\ell$. In this section we shall consider the unequal case.

Consider a partition of the points in EG(N,q) into flat spaces, say $\pi_1, \pi_2, \ldots, \pi_\ell$. Then

$$\pi_i \cap \pi_j = \emptyset , \quad \text{for all } i \text{ and } j$$

and the join of $\pi_1, \pi_2, \ldots, \pi_\ell$ is the whole space. Also $\pi_1, \pi_2, \ldots, \pi_\ell$ are not necessarily of the same dimension. Notice that every point of EG(N,q) belongs to exactly one flat space $\pi_i$ $(i = 1,2,\ldots,\ell)$. These flat spaces are identified with the attributes. So the number of levels of any attribute will be of the form

$$n_i = q^{h_i} , \quad 0 \le h_i < N, \quad \text{for all } i.$$

If $h_1 = h_2 = \cdots = h_\ell$ then $\pi_1, \pi_2, \ldots, \pi_\ell$ form a parallel bundle in EG(N,q). This case is studied by Ghosh and Abraham [23].

Example (2.6.1). As an illustration, consider the case $\ell = 3$, $n_1 = 4$, $n_2 = n_3 = 2$. The filing scheme for these attributes can be constructed using an EG(3,2). The points of the geometry are triplets and for simplicity we shall write their coordinates without commas i.e. $(x_1x_2x_3)$. Thus the points are

000, 001, 010, 011
100, 101, 110, 111.

The three flats $\pi_1$, $\pi_2$ and $\pi_3$ are

$\pi_1$: $x_1 = 0$ : 000, 001, 010, 011
$\pi_2$: $x_1=1, x_2=0$: 100, 101
$\pi_3$: $x_1=1, x_2=1$: 110, 111.

These flat spaces are constructed as follows: first a parallel bundle of order one is formed i.e. $\{x_1=0\}$; $\{x_1=1\}$; then a parallel bundle of order one in $\{x_1=1\}$ is found i.e. $\{x_1=1;\ x_2=0\}$; $\{x_1=1,\ x_2=1\}$. The correspondence between points and attribute levels is as follows:

$A_1 = \pi_1$ so $A_{11} = 000$, $A_{13} = 010$,
$A_{12} = 001$, $A_{14} = 011$.

$A_2 = \pi_2$ so $A_{21} = 100$, $A_{22} = 101$.
$A_3 = \pi_3$ so $A_{31} = 110$, $A_{32} = 111$.

The buckets are constructed by identifying them with planes, not containing $\pi_1$ or $\pi_2$ or $\pi_3$. So

| Bucket No. | Identification | Equation | Attribute levels |
|---|---|---|---|
| 1 | 0010 | $x_3=0$ | $A_{11}A_{13}A_{21}A_{31}$ |
| 2 | 0011 | $x_3=1$ | $A_{12}A_{14}A_{22}A_{32}$ |
| 3 | 1010 | $x_1+x_3=0$ | $A_{11}A_{13}A_{22}A_{32}$ |
| 4 | 1011 | $x_1+x_3=1$ | $A_{12}A_{14}A_{21}A_{31}$ |
| 5 | 0110 | $x_2+x_3=0$ | $A_{11}A_{14}A_{21}A_{32}$ |
| 6 | 0111 | $x_2+x_3=1$ | $A_{12}A_{13}A_{22}A_{31}$ |
| 7 | 1110 | $x_1+x_2+x_3=0$ | $A_{11}A_{14}A_{22}A_{31}$ |
| 8 | 1111 | $x_1+x_2+x_3=1$ | $A_{12}A_{13}A_{21}A_{32}$ |

The identification number attached to each bucket is the 1-tuple of the coefficients of the equation $\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = 0$ (where $\lambda_i \in GF(2)$) of the plane corresponding to the bucket. Within each bucket, the accession numbers of the records are divided into subsets called subbuckets, corresponding to each relevant triplet of values. The subbuckets may be identified by concatenating the codes of the triplet of values they represent. As an example,

| Bucket Identification | Subbucket Identification | Accession nos. |
|---|---|---|
| 0010 | 000 100 110 | ... |
| | 010 100 110 | ... |
| 0011 | 001 100 110 | ... |
| | 011 100 110 | ... |
| 1010 | 000 101 111 | ... |
| | 010 101 111 | ... |
| 1011 | 001 100 110 | ... |
| | 011 100 110 | ... |
| 0110 | 000 100 111 | ... |
| | 011 100 111 | ... |
| 0111 | 001 101 110 | ... |
| | 010 101 110 | ... |
| 1110 | 000 101 110 | ... |
| | 011 101 110 | ... |
| 1111 | 001 100 111 | ... |
| | 010 100 111 | ... |

The buckets corresponding to $x_2 = 0$; $x_2 = 1$; $x_1+x_2 = 0$; $x_1+x_2 = 1$ are deleted from the filing scheme, as they do not contain any relevant triplets. The remaining eight flats intersect all the three flats, $\pi_1$, $\pi_2$ and $\pi_3$.

Suppose the query request is to retrieve all records pertaining to $A_{11}$, $A_{21}$ and $A_{31}$. Then $A_{11}$, $A_{21}$ and $A_{31}$ are first converted into the points of the geometry. These points are (000), (100), (110). The plane passing through these three points is determined by solving the equation

$$\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = 0$$

in GF(2). On substituting these points in the equation, we have

$$\lambda_0 = 0, \quad \lambda_1 = 0.$$

and

$$\lambda_1 + \lambda_2 = 0 \quad \Rightarrow \quad \lambda_2 = 0$$

Hence the equation is $x_3 = 0$. So the bucket containing these three values is 0010. The accession numbers corresponding to the query are obtained from the subbucket 000 100 110 of the bucket 0010.

The retrieval time for a query A is the sum of three components:

$$T(A) = T_1 + T_2 + T_3$$

where $T_1$ = the time needed to solve the algebraic equation,

$T_2$ = the time required for matching identification, numbers of the buckets

$T_3$ = the time required for identifying the subbuckets.

# CHAPTER III

# QUADRICS AND SOME RELATED FILING SCHEMES

In this chapter we shall describe a method for obtaining filing schemes based on the properties of a homogeneous, non-degenerate quadric in a finite projective geometry. Before describing these procedures we shall describe some important properties of quadrics.

## 3.1 Quadrics

The following remarks are based on the work of Primrose [27], Ray-Chaudhuri [30] and Bose [6]. Our introductory comments are those of Dowling [21].

Let $\sum$ denote a projective geometry of order $N$ over $GF(q)$. Let $B$ be an $(N+1) \times (N+1)$ matrix with elements from $GF(q)$ and consider the equation

$$\underline{x}\, B\, \underline{x}^T = 0 \tag{3.1.1}$$

where $\underline{x} = (x_0,\ldots,x_N)$ is a row vector with elements $x_i \in GF(q)$, $i = 0,1,2,\ldots,N$ and $\underline{x}^T$ is the transpose of $\underline{x}$. The set of points in $PG(N,q)$ whose coordinate vectors satisfy (3.1.1) is said to constitute a *quadric* in $PG(N,q)$. Without loss of generality we can take $B$ to be an upper triangular matrix, i.e.

$$B = \begin{bmatrix} b_{00} & b_{01} & b_{02} & \cdots & b_{0N} \\ 0 & b_{11} & b_{12} & \cdots & b_{1N} \\ 0 & 0 & b_{22} & \cdots & b_{2N} \\ & & & & \\ 0 & 0 & 0 & \cdots & b_{NN} \end{bmatrix} \tag{3.1.2}$$

for if B is an arbitrary $(N+1) \times (N+1)$ matrix with elements from $GF(q)$, then clearly the matrix D defined by

$$
\begin{aligned}
b_{ij} &= 0 , \quad i > j, \\
b_{ii} &= d_{ii}, \\
b_{ij} &= d_{ij} + d_{j2} \quad i < j,
\end{aligned}
$$

is upper triangular, and

$$
\begin{aligned}
\underline{x}\, D\, \underline{x}^{T} &= \sum_{i=0}^{N} \sum_{j=0}^{N} d_{ij}\, x_i x_j \\
&= \sum_{0 \le i \le j \le N} b_{ij}\, x_i x_j \\
&= \underline{x}\, B\, \underline{x}^{T}.
\end{aligned}
$$

As both the equations $\underline{x}\, B\, \underline{x}^{T} = 0$ and $\underline{x}\, D\, \underline{x}^{T} = 0$ represent the same quadric, we take the equation

$$\underline{x}\, B\, \underline{x}^{T} = 0$$

(i.e. (3.1.1)) to represent the quadric.

Now let

$$Q_N\colon \ \underline{x}\, B\, \underline{x}^{T} = 0$$

be a quadric in $PG(N,q)$, and let D be a non-singular $(N+1) \times (N+1)$ matrix with elements from $GF(q)$. The non-singular linear transformation

$$\underline{x}^{*T} = D^{-1}\, \underline{x}^{T}$$

carries the point P with row vector $\underline{x}$ into the point P* with row vector $\underline{x}^{*}$ and the quadric $Q_N$ into the quadric

$$Q_N^{*}\colon \ \underline{x}^{*}\, B^{*}\, \underline{x}^{*T} = 0 ,$$

where $B^{*}$ is the triangular matrix obtained from $C = DBD^{T}$ in the manner described above. This transformation is incidence preserving, i.e. $P^* \in Q_N^*$ if and only if $P \in Q_N$. The quadrics $Q_N$ and $Q_N^*$ are called *equivalent*. If $t$ is the largest integer for which there exists a non-singular matrix $D$ such that the last $t$ columns of the corresponding $B^*$ are null, then the *rank* of $Q_N$ and all equivalent quadrics is defined to be $N + 1 - t$. If $t = 0$, then $Q_N$ is said to be *non-degenerate;* otherwise $Q_N$ is *degenerate*.

Thus if the rank of $Q_N$ is $N + 1 - t$ where $t \geq 1$, then by an appropriate linear transformation $\underline{x}^{*T} = D^{-1}\, \underline{x}^T$ we can find an equivalent quadric

$$Q^*_N: \sum_{0\leq i\leq j\leq N-t} b^*_{ij}\, x^*_i\, x^*_j = 0 . \tag{3.1.3}$$

Notice that in equation (3.1.3), the variables $x^*_{N-t+1}, x^*_{N-t+2}, \ldots, x^*_N$ are missing. Define the $(N-t)$-flat

$$\sum\nolimits^*_{N-t} : \ x^*_{N-t+1} = x^*_{N-t+2} = \cdots = x^*_N = 0 . \tag{3.1.4}$$

Then since the rank of $Q^*_N$ is $N+1-t$, the quadric

$$Q^*_{N-t}: \sum_{0\leq i\leq j\leq N-t} b^*_{ij}\, x^*_i\, x^*_j = 0,\ x^*_{N-t+1}=0,\ldots,x^*_N = 0 \tag{3.1.5}$$

is clearly non-degenerate in $\sum^*_{N-t}$. We can write

$$Q^*_{N-t} = Q^*_N \cap \sum\nolimits^*_{N-t} .$$

If $\sum_{N-t}$ is the inverse image of $\sum^*_{N-t}$ under the transformation $\underline{x}^{*T} = D^{-1}\, \underline{x}^T$, then the quadric

$$Q_{N-t} = \sum\nolimits_{N-t} \cap Q_N$$

is non-degenerate in $\sum_{N-t}$.

The equivalent quadrics $Q_N$ and $Q^*_N$ are called *cones of order* $t$ in $PG(N,q)$. If $P^*_0$ is a point of the $(t-1)$-flat

$$\sum^*_{t-1}:\ x^*_0 = 0,\ x^*_1 = 0,\ \ldots,\ x^*_{N-t} = 0$$

and $P^*_1$ is a point of $Q^*_{N-t}$, then clearly any point $P^*$ on the line $(P^*_0\, P^*_1)$ lies in $Q^*_N$. Conversely, if $P^* \in Q^*_N$, $P^* \in (P^*_0\, P^*_1)$ for some $P^*_0 \in \sum^*_{t-1}$, $P^*_1 \in Q^*_{N-t}$. The (t-1)-flat $\sum^*_{t-1}$ is called the *vertex* of the cone $Q^*_N$, and the non-degenerate quadric $Q^*_{N-t}$ is called the *base* (of course the base is not unique). If $\sum_{t-1}$ is the inverse image of $\sum^*_{t-1}$ under the transformation $\underline{x}^{*T} = D^{-1}\underline{x}^T$ then $\sum_{t-1}$ is the vertex and $Q_{N-t}$ is the base of the cone $Q_N$. Unless otherwise stated, we shall employ the term *cone* to refer to a cone of order one. In this case, the vertex consists of a single point and the base in a non-degenerate quadric in (N-1) dimensions.

If $Q_N$ is a non-degenerate quadric in PG(N,q), where $N = 2k$ is even then $Q_N$ contains flat spaces of dimension $k - 1$ but not of any higher dimension. But if $N = 2K+1$ (i.e. odd), the non-degenerate quadrics are of two types. An *elliptic quadric* in PG(2k+1, q) contains flat spaces of dimension $k - 1$ but none of any higher dimension. A *hyperbolic quadric* in PG(2k+1, q) contains flat spaces of dimension k but none of any higher dimension.

Let us denote by $\psi(N,0)$ the number of points in a non-degenerate quadric $Q_N$ in PG(N,q). Primrose [27] showed that

$$\psi(N,0) = \begin{cases} f_0(k,0) & \text{if } N = 2k, \\ f_1(k,0) & \text{if } N = 2k+1 \ \&\ Q_N \text{ is elliptic,} \\ f_2(k,0) & \text{if } N = 2k+1 \ \&\ Q_N \text{ is hyperbolic,} \end{cases}$$

where

$$f_0(k,0) = (q^{2k}-1)/(q-1)\,, \tag{3.1.6}$$

$$f_1(k,0) = (q^{k+1}+1)(q^k-1)/(q-1)\,, \tag{3.1.7}$$

$$f_2(k,0) = (q^{k+1}-1)(q^k+1)/(q-1)\,. \tag{3.1.8}$$

Ray-Chaudhuri [30] generalized this result by finding the number, $\psi(N,r)$, of r-flats contained in a non-degenerate quadric $Q_N$ of $PG(N,q)$. He showed that

$$\psi(N,r)=\begin{cases} f_0(k,r) & \text{if } N = 2k\,, \\ f_1(k,r) & \text{if } N = 2k+1 \ \&\ Q_N \text{ elliptic}\,, \\ f_2(k,r) & \text{if } N = 2k+1 \ \&\ Q_N \text{ hyperbolic}, \end{cases}$$

where

$$f_0(k,r) = \begin{cases} \prod_{m=0}^{r} (q^{2k-2m}-1)/(q^{r-m+1}-1) & \text{if } r \le k-1 \\ 0 & \text{if } r > k-1 \end{cases} \tag{3.1.9}$$

$$f_1(k,r) = \begin{cases} \prod_{m=0}^{r} (q^{k-m+1}+1)(q^{k-m}-1)/(q^{r-m+1}-1) & \text{if } r \le k-1 \\ 0 & \text{if } r > k-1 \end{cases} \tag{3.1.10}$$

$$f_2(k,r) = \begin{cases} \prod_{m=0}^{r} (q^{k-m+1}-1)(q^{k-m}+1)/(q^{r-m+1}-1) & \text{if } r \le k \\ 0 & \text{if } r > k\,. \end{cases} \tag{3.1.11}$$

If $Q_N$ is a cone in PG(N,q), where N is even, then $Q_N$ is said to be an elliptic or a hyperbolic cone according as its base $Q_{N-1}$ is elliptic or hyperbolic. If N is odd, then $Q_N$ will simply be called a cone. In general if r is the dimension of the highest-dimensional-flat space contained in the base $Q_{N-1}$ of a cone $Q_N$, then $Q_N$ contains (r+1)-flats but none of any higher dimension. Thus when N = 2k, an elliptic cone contains (k-1)-flats and a hyperbolic cone contains k-flats.. If N = 2k+1, a cone contains k-flats also. If $\psi(N-1, 0)$ is the number of points contained in the base $Q_{N-1}$ of a cone $Q_N$, then the number of points in $Q_N$ is

$$1 + q\,\psi(N-1, 0). \tag{3.1.12}$$

By a suitable choice of the system of reference, the equation of any non-degenerate quadric $Q_N$ in PG(N,q) can be expressed in a relatively simple form. If $Q_N$ is a non-degenerate quadric in PG(N,q), where N = 2k, then the equation of $Q_N$ can be expressed in the canonical form

$$Q_N:\ x_0x_1 + x_2x_3 + \ldots + x_{2k-2}x_{2k-1} + x_{2k}^2 = 0. \tag{3.1.13}$$

The equation of a non-degenerate hyperbolic quadric $Q_N$ in PG(N,q), where N = 2k+1, can be written as

$$Q_N:\ x_0x_1 + x_2x_3 + \ldots + x_{2k}x_{2k+1} = 0. \tag{3.1.14}$$

If $q = p^n$ where p a prime number not equal to two, and n is an integer, then the equation of a non-degenerate elliptic quadric $Q_N$ in PG(N,q), (N=2k+1), can be written as

$$Q_N:\ x_0x_1 + x_2x_3 + \ldots + x_{2k-2}x_{2k-1} + x_{2k}^2 + \beta x_{2k+1}^2 = 0, \tag{3.1.15}$$

where "$-\beta$" is a non-square element of $GF(q)$. If $p = 2$, the equation of the elliptic quadric $Q_N$, $(N=2k+1)$, can be written as

$$Q_N: \; x_0x_1 + x_2x_3 + \dots + x_{2k-2}x_{2k-1} + \lambda(x_{2k}^2 + x_{2k+1}^2) + x_{2k}x_{2k+1} = 0, \qquad (3.1.16)$$

where $\lambda(x_{2k}^2 + x_{2k+1}^2) + x_{2k}x_{2k+1}$ is irreducible over $GF(2^n)$.

Earlier we defined the rank of a quadric $Q_N$ in $PG(N,q)$. If $q$ is odd (i.e. $p \neq 2$) then the rank of $Q_N$ is the same as the rank of the symmetric matrix $B + B^T$, and the equation of $Q_N$ may be taken as

$$\underline{x}(B+B^T)\,\underline{x}^T = 0.$$

However, this is not generally true for fields of characteristic two (2). If $Q_N$ is a non-degenerate quadric in $PG(N,2^n)$, then the rank of $B + B^T$ is $N + 1$ (i.e. $B + B^T$ is non-singular as in the case $p \neq 2$) if $N$ is odd, but $B + B^T$ is singular of rank $N$ when $N$ is even. Conversely, if at least one of $N$ and $q$ is odd, $Q_N$ is non-degenerate if $B + B^T$ is non-singular. If both $N$ and $q$ are even, $Q_N$ is non-degenerate if the rank of $B + B^T$ is $N$.

The matrix $B + B^T$ defines a *polarity* with respect to the non-degenerate quadric $Q_N$ according to which the points $P_0$ and $P_1$ with coordinate vectors $\underline{P}_0$ and $\underline{P}_1$ are said to be *conjugate* if and only if

$$\underline{P}_0(B + B^T)\underline{P}_1^T = 0.$$

The relation of conjugacy is obviously symmetrical. If $q$ is even, then every point of $PG(N,q)$ is self conjugate. If $q$ is odd then the only self conjugate points are those of $Q_N$.

A point $P$ with row vector $\underline{P}$ is said to be a *regular point* with respect to $Q_N$ if

$$\underline{P}(B + B^T) \neq 0;$$

otherwise P is said to be an *irregular point.* Since $B + B^T$ is non-singular when at least one of N and q is odd, every point is regualr in these cases. If both N and q are even, then $B + B^T$ is singular of rank N and there exists a unique point C with row vector $\underline{C}$ such that

$$\underline{C}(B + B^T) = 0.$$

The irregular point C is called the *nucleus of polarity* of the quadric $Q_N$. All other points of PG(N,q) are regular.

The *polar space* T(P) of a point P is defined to be the set of all points conjugate to P, that is

$$T(P){:}\ \ \underline{P}(B + B^T)\underline{x}^T = 0 \tag{3.1.17}$$

where $\underline{P}$ is the row vector corresponding to the point P. If P is a regular point, then (3.1.17) is the equation of an (N-1)-flat. If P is the nucleus of polarity, however, then every point is conjugate to P and hence T(P) is the entire space. Since the relation of conjugacy is symmetrical, for any two points $P_0$, $P_1$ we have $P_0 \in T(P_1)$ if and only if $P_1 \in T(P_0)$. It follows that the polar space of every point $PG(2k, 2^n)$ passes through the nucleus of polarity C. If $P_0$ is a point distinct from C and $P_1$ is any other point on the line $(P_0C)$ then the row vector $\underline{P}_1$ for $P_1$ has the form

$$\underline{P}_1 = \underline{C} + \lambda\ P_0,$$

where $\lambda \neq 0$ and $\underline{P}_0$ is the row vector of $P_0$. Then

$$\begin{aligned}\underline{P}_1(B + B^T)\underline{x}^T &= (\underline{C} + \lambda\underline{P}_0)(B + B^T)\underline{x}^T \\ &= \lambda\ \underline{P}_0(B + B^T)\underline{x}^T\ .\end{aligned}$$

So it follows that

$$P_1(B + B^T)\underline{x}^T = 0$$

if and only if

$$\underline{P}_0(B + B^T)\underline{x}^T = 0.$$

So

$$T(P_1) = T(P_0).$$

If, however, at least one of $N$ and $q$ is odd, then every point is regular and hence there is a one-one correspondence between points and their polar spaces.

The polar space $T(\sum_t)$ of a t-flat, $\sum_t$, is the set of all points conjugate to every point of $\sum_t$. If $P_0, P_1, \ldots, P_t$ are $(t + 1)$ independent points of $\sum_t$, then $T(\sum_t)$ is the $(N-t-1)$-flat

$$T(\textstyle\sum_t): \ \underline{P}_i(B + B^T)\underline{x}^T = 0, \quad i = 0,1,2,\ldots,t,$$

where $\underline{P}_i$ is the row vector of $P_i$. $T(\sum_t)$ is clearly the intersection of the $(N-1)$-flats $T(P_i)$, $i = 0,1,2,\ldots,t$. If $\sum_t$ and $\sum_r$ are any two flat spaces, then $\sum_t \subset T(\sum_r)$ if and only if $\sum_r \subset T(\sum_t)$. A t-flat $\sum_t$ is contained in $Q_N$ if and only if the points $P_0, P_1, P_2, \ldots, P_t$ are pairwise conjugate. Thus if $\sum_t \subset Q_N$, then $\sum_t \subset T(\sum_T)$.

## 3.2 (N-1)-flat spaces and quadrics

Theorem (3.2.1)[†]. Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$ and let $P_0$ be a point on $Q_N$. Then the quadric $Q_{N-1} = Q_N \cap T(P_0)$ is a cone in $T(P_0)$ with vertex $P_0$.

---

[†] This is a well known result. The proof given here, is that of Dowling [21].

Proof. Let $P_2, P_3, \ldots, P_N$ be $(N-1)$ independent points of an $(N-2)$-flat $\sum_{N-2}$, where $\sum_{N-2} \subset T(P_0)$ and $\sum_{N-2}$ does not contain $P_0$. Further let $P_1 \notin T(P_0)$. Since $P_0 \in Q_N$, $P_0 \in T(P_0)$. Hence the points $P_i (i=0,1,2,\ldots,N)$ are independent, and we can take them, without loss of generality, to be unit points. So the row vector of the point $P_i$ is

$$\underline{P}_i = (\underbrace{0,0,\ldots,0}_{i},1,\underbrace{0,\ldots,0}_{N-i}), \quad i = 0,1,2,\ldots,N. \qquad (3.2.1)$$

We can then write

$$\textstyle\sum_{N-2}: \quad x_0 = 0, \qquad x_1 = 0 \qquad (3.2.2)$$

$$T(P_0): x_1 = 0 \qquad (3.2.3)$$

But, if $\underline{x}\, B\, \underline{x}^T$ is the equation of the quadric $Q_N$ then

$$T(P_0): \quad \underline{P}_0(B + B^T)\underline{x}^T = 0. \qquad (3.2.4)$$

Comparing (3.2.4) and (3.2.3) we have

$$b_{00} = b_{02} = b_{03} = \ldots = b_{0N} = 0$$

and

$$b_{01} \neq 0$$

Hence, we have

$$Q_N: \quad x_1(b_{01}x_0 + b_{11}x_1 + b_{12}x_2 + \ldots + b_{1N}x_N) + Q(x_2,\ldots,x_N) = 0, \qquad (3.2.5)$$

where

$$Q(x_2,x_3,\ldots,x_N) = \sum_{2\le i\le j\le N} b_{ij}x_ix_j .$$

It follows from (3.2.3) and (3.2.5) that

$$Q_{N-1} = Q_N \cap T(P_0): \quad x_1 = 0 , \quad Q(x_2,\ldots,x_N) = 0 \qquad (3.2.6)$$

Now since $Q_N$ is non-degenerate it is clear from equation (3.2.5) that the quadric

$$Q_{N-2} = \Sigma_{N-2} \cap Q_N: \quad x_0 = 0,\ x_1 = 0,\ Q(x_2,x_3,\ldots,x_N) = 0 \qquad (3.2.7)$$

is non-degenerate in $\Sigma_{N-2}$. Hence $T(P_0) \cap Q_N$ is a cone in $T(P_0)$. Clearly the vertex of the cone is $P_0$.

<u>Lemma (3.2.2)</u>. Let $Q_{2k}$ be a non-degenerate quadric in $PG(2k,2^n)$, $n$ an integer. Let $C$ be the nucleus of polarity. Then corresponding to each $P \in Q_N$, there exists a $(2k-1)$-flat, the polar space of $P$, $T(P)$, such that $C$ is a point of $T(P)$. Further the polar spaces corresponding to the points of $Q_N$ are all distinct.

<u>Proof</u>. Let $\underline{C}$ be the row vector corresponding to $C$, the nucleus of polarity. Since $Q_{2k}$ is non-degenerate and $q - 2^n$, we can take without loss of generality $Q_{2k}$ as

$$x_0x_1 + x_2x_3 + \ldots + x_{2k-2}x_{2k-1} + x_{2k}^2 = 0 \qquad (3.2.8)$$

i.e.

$$\underline{x}\, B\, \underline{x}^T = 0$$

where

$$\underset{(2k+1)\times(2k+1)}{B} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix} \qquad (3.2.9)$$

Hence, $\underline{C} = (0,0,0,\dots,0,0,1)$. Let $P$ be any point in $Q_N$. Then its polar space $T(P)$ is

$$\underline{P}(B + B^T)\underline{x}^T = 0, \tag{3.2.10}$$

where $\underline{P}$ is the row vector of $P$. Since $C$ is the nucleus of polarity, we have

$$\underline{C}(B + B^T) = 0. \tag{3.2.11}$$

Hence,

$$C \in T(P).$$

Let $P_1$ and $P_2$ be any two distinct points in $Q_N$. Then their polar spaces are

$$\underline{P}_1 = (B + B^T)\underline{x}^T = 0, \tag{3.2.12}$$

$$\underline{P}_2 = (B + B^T)\underline{x}^T = 0. \tag{3.2.13}$$

If $T(P_1) = T(P_2)$ then the solution spaces of the equations (3.2.12) and (3.2.13) are equivalent. Hence

$$\underline{P}_1 = \lambda\, \underline{P}_2,$$

where $\lambda$ is a scalar belonging to $GF(2^n)$. That is $\underline{P}_1$ and $\underline{P}_2$ represent the same point, i.e. $P_1 = P_2$. But this is a contradiction as $P_1$ and $P_2$ are two distinct points. Hence $T(P_1) \neq T(P_2)$. So the result.

Lemma (3.2.3). If $\sum_{2k-1}$ is any $(2k-1)$-flat in $PG(2k,2^n)$ such that it is not the polar space of any point in the geometry, then the nucleus of polarity does not belong to $\sum_{2k-1}$.

Proof. Without loss of generality we can take $Q_{2k}$ as

$$x_0x_1 + x_2x_3 + \dots + x_{2k-2}x_{2k-1} + x_{2k}^2 = 0.$$

Then $\underline{C}$ is $(0,0,0,0,\dots,0,1)$. Now it is easy to notice that a (2k-1)-flat will not contain the nucleus of polarity, C if and only if the linear equation representing the (2k-1)-flat does not involve the variable $x_{2k}$. Further, any equation with $x_{2k}$, cannot be written as

$$\underline{P}(B + B^T)\underline{x}^T = 0,$$

for some point P in $PG(2k,2^n)$. Hence the result.

So we can conclude that there exists a one-one correspondence between points of $Q_{2k}$ and (2k-1)-flats containing the nucleus of polarity. Further the rest of the (2k-1)-flats, not containing the nucleus of polarity cannot be expressed as polar spaces of points in $PG(2k,2^n)$.

Theorem (3.2.4)[†]. If $\Sigma_{2k-1}$ is any (2k-1)-flat in $PG(2k,2^n)$ not containing the nucleus of polarity C, then $Q_{2k-1} = \Sigma_{2k-1} \cap Q_{2k}$ is non-degenerate.

Proof. As before, let us take $Q_N$ as

$$x_0x_1+x_2x_3+\dots+x_{2k-2}x_{2k-1}+x_{2k}^2 = 0,$$

and so the row vector of the nucleus of polarity C is $(0,0,0,\dots,0.1)$. By lemma (3.2.3) we have the equation of $\Sigma_{2k-1}$ as

$$L(x_0,x_1,\dots,x_{2k-1}) + x_{2k} = 0 \qquad (3.2.14)$$

where $L(x_0,x_1,\dots,x_{2k-1})$ is a linear form in $x_0,x_1,\dots,x_{2k-1}$. So the quadric $Q_{2k-1}$ can be written as

$$x_0x_1+x_2x_3+\dots+x_{2k-2}x_{2k-1}+\{L(x_0,x_1,\dots,x_{2k-1})\}^2=0. \qquad (3.2.15)$$

---

[†] Dowling [22] has shown that, in $PG(2k,2)$, $Q_{2k-1}$ is non-degenerate if and only if $C \notin \Sigma_{2k-1}$.

Since our base field is $GF(2^n)$, we will not have any cross products in the expansion of $\{L(x_0,x_1,\dots,x_{2k-1})\}^2$. So the $(2k \times 2k)$ matrix $G$ of $Q_{2k-1}$ will be

$$\underset{2k \times 2k}{G} = \begin{bmatrix} g_{00} & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & g_{11} & 0 & 0 & \dots & 0 & 0 \\ & & & & & & \\ 0 & 0 & 0 & 0 & \dots & g_{2k-2,2k-2} & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & g_{2k-1,2k-1} \end{bmatrix} \quad (3.2.15)$$

where $g_{ii}$ is either 1 or 0 according as $x_i$ occurs in the linear form $L(x_0,x_1,\dots,x_{2k-1})$ or not. It is easy to notice that

$$\text{rank}(G + G^T) = \text{rank}(B + B^T) = 2k,$$

where $B$ is the matrix of $Q_{2k}$. Hence $Q_{2k-1}$ is non-degenerate in $\sum_{2k-1}$.

Theorem (3.2.5)[†]. Let $Q_{2k}$ be a non-degenerate quadric in $PG(2k,q=2^n)$. Define $Q_{2k-1} = Q_{2k} \cap \sum_{2k-1}$, where $\sum_{2k-1}$ is a (2k-1)-flat. Then there are $N_1$, (2k-1)-flats such that $Q_{2k-1}$ is a cone of order one; $N_2$, (2k-1)-flats such that $Q_{2k-1}$ is a non-degenerate hyperbolic quadric and $N_3$, (2k-1)-flats such that $Q_{2k-1}$ is a non-degenerate elliptic quadric; where

$$N_1 = (q^{2k}-1)/(q-1),$$
$$N_2 = (q^{2k} + q^k)/2,$$

---

[†] This result when $n = 1$, is obtained by Dowling [22].

and

$$N_3 = (q^{2k} - q^k)/2 \qquad , \quad q = 2^n .$$

Proof. To determine $N_t$ $(t = 1,2,3)$ we count the number of pairs $(P, \Sigma_{2k-1})$ where $P$ is a point of $Q_{2k}$ and $\Sigma_{2k-1}$ is a hyperplane containing $P$. Since, each point is contained in $(q^{2k}-1) \div (q-1)$ hyperplanes and there are $(q^{2k}-1)/(q-1)$ points in $Q_{2k}$, the number of pairs is $(\frac{q^{2k}-1}{q-1})^2$. Counting these pairs in another way,

$$\left(\frac{q^{2k}-1}{q-1}\right)^2 = N_1\left(\frac{q^{2k}-1}{q-1}\right) + N_2\left(\frac{q^{2k-1}-1}{q-1} + q^{k-1}\right) + N_3\left(\frac{q^{2k-1}-1}{q-1} - q^{k-1}\right) \tag{3.2.16}$$

i.e.

$$\frac{q^{4k}-2q^{2k}+1}{(q-1)^2} = (N_1+N_2+N_3) \times \left(\frac{q^{2k-1}-1}{q-1}\right) - (N_3-N_2)q^{k-1} .$$

But

$$N_1 + N_2 + N_3 = \frac{q^{2k+1}-1}{q-1} . \tag{3.2.17}$$

From theorem (3.2.4)

$$N_1 = \left(\frac{q^{2k}-1}{q-1}\right) . \tag{3.2.18}$$

So

$$N_3 + N_2 = q^{2k} . \tag{3.2.19}$$

From (3.2.17) and (3.2.16) we have

$$N_2 - N_3 = q^k . \tag{3.2.20}$$

Hence

$$N_2 = (q^{2k} + q^k)/2 ,$$

and

$$N_3 = (q^{2k} - q^k)/2 .$$

$N_1$, $N_2$ and $N_3$ are integers as $q = 2^n$.

Theorem (3.2.6). Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$, $N \geq 2$, where at least one of $N$ and $q$ is odd. Then if $P_0 \notin Q_N$ the quadric $Q_{N-1} = T(P_0) \cap Q_N$ is non-degenerate in $T(P_0)$.

This result is obtained by Dowling [21]. We will not give the proof of this theorem.

Theorem (3.2.7). Let $Q_{2k+1}$ be a non-degenerate hyperbolic quadric in $PG(2k+1, q)$. Then there are $(q^{2k+1}-q^k)$ 2k-flat spaces that intersect $Q_{2k+1}$ in a non-degenerate quadric and $(q^{k+1}-1)(q^k+1)/(q-1)$ 2k-flat spaces that intersect in hyperbolic cones of order one.

Proof. From theorem (3.2.1) and theorem (3.2.6), as there is one-one correspondence between (2k-1)-flat spaces and polar spaces with respect to $Q_{2k+1}$, we have $(q^{2k+1}-q^k)$ 2k-flat spaces intersecting $Q_{2k+1}$ in a non-degenerate quadric and $(q^{k+1}-1)(q^k+1)/(q-1)$ 2k-flat spaces that intersect in a cone of order one. These cones can be elliptic cones or hyperbolic cones.

Suppose one of these cones is an elliptic cone. Let us denote it by $Q^*_{2k}$. Then the base of $Q^*_{2k}$, say $Q^*_{2k-1}$, is non-degenerate in a (2k-1)-flat space and is elliptic. Hence $Q^*_{2k-1}$ contains (k-2)-flat spaces but no higher flat spaces. That is, $Q_{2k+1}$ will contain only (k-1)-flat spaces but no higher. But this is a contradiction as $Q_{2k+1}$ is a hyperbolic cone. Hence all cones are hyperbolic cones.

Theorem (3.2.8). Let $Q_{2k+1}$ be an elliptic quadric in $PG(2k+1, q)$. Then there are $(q^{2k+1} + q^k)$ 2k-flat spaces that intersect $Q_{2k+1}$ in a non-degenerate quadric and $(q^{k+1}+1)(q^k-1)/(q-1)$ 2k-flat spaces that intersect in an elliptic cone of order one.

**Proof.** The first part of the theorem follows from theorem (3.2.6) and theorem (3.2.1). That is, there are $(q^{k+1}-1)(q^k-1)/(q-1)$ cones of order one. Suppose one of these cones is hyperbolic. Then the base of this cone will contain (k-1)-flat spaces but no higher. Hence $Q_{2k+1}$ will contain k-flat spaces. But $Q_{2k+1}$ is elliptic and hence it will not contain k-flat spaces. This is a contradiction. Hence all cones are elliptic cones.

## 3.3 (N-2)-flat spaces and quadrics

**Lemma (3.3.1)**[†]. The number of points in a degenerate quadric of order r in PG(n,q), is

$$\frac{q^r-1}{q-1} + q^r \, |Q_{N-r}| \, . \tag{3.3.1}$$

**Proof.** Since $Q_N$ is a cone of order r, there exists a unique (r-1)-flat $\sum_{r-1}$, called the vertex of $Q_N$, such that the points of $Q_N$ are those of the lines joining, points of $\sum_{r-1}$ to the points of $Q_{N-r}$; where $Q_{N-r}$ is a non-degenerate quadric in (N-r)-dimensions obtained by intersecting $Q_N$ with any (N-r)-flat, $\sum_{N-r}$, which is skew to $\sum_{r-1}$.

Hence

$$\begin{aligned} |Q_N| &= |\textstyle\sum_{r-1}| + (q-1)|\textstyle\sum_{r-1}| \times |Q_{N-r}| + |Q_{N-r}| \\ &= \frac{q^r-1}{q-1} + (q-1) \times \frac{(q^r-1)}{(q-1)} \times |Q_{N-r}| + |Q_{N-r}| \\ &= \frac{q^r-1}{q-1} + q^r \times |Q_{N-r}| \, . \end{aligned}$$

---

[†] This is a well known result. Reference Bose [6].

Theorem (3.3.2). Let $Q_N$ be a cone of order $r$ in $PG(N,q=2^n)$, where $(N-r)$ is even. Let $\sum_r$ be the nucleus of polarity of $Q_N$ and $\sum_{r-1} = \sum_r \cap Q_N$ the vertex. Let $\sum_{N-1}$ be a hyperplane and define $Q_{N-1} = Q_N \cap \sum_{N-1}$. Then if $\sum_{r-1} \not\subset \sum_{N-1}$

1. $Q_{N-1}$ is a cone of order $r-1$;

if $\sum_r \subset \sum_{N-1}$,

2. $Q_{N-1}$ is a cone of order $r+1$;

if $\sum_{r-1} \subset \sum_{N-1}$ but $\sum_r \not\subset \sum_{N-1}$, then either

3. $Q_{N-1}$ is an hyperbolic cone of order $r$,

or

4. $Q_{N-1}$ is an elliptic cone of order $r$.

The number $N_t$, $t = 1,2,3,4$ of hyperplanes $\sum_{N-1}$ for which $Q_{N-1}$ is of type $t$ is

$$N_1 = q^{N-r+1}(q^r-1)/(q-1) ,$$
$$N_2 = (q^{N-r}-1)/(q-1) ,$$
$$N_3 = (q^{N-r}+q^{(N-r)/2})/2 ,$$

and

$$N_4 = (q^{N-r}+q^{(N-r)/2})/2 . \qquad (3.3.2)$$

Proof[†]. If $\sum_{r-1} \not\subset \sum_{N-1}$, then $\sum_{N-1}$ contains $(N-r)$-flat $\sum_{N-r}$ skew to $\sum_{r-1}$. The quadric $Q_{N-r} = Q_N \cap \sum_{N-r}$ is non-degenerate in $\sum_{N-r}$, and $Q_{N-1}$ clearly consists of all points on the lines joining points of $Q_{N-r}$ to points of $\sum_{r-2} = \sum_{r-1} \cap \sum_{N-1}$. Hence, $Q_{N-1}$ is a cone of order $r-1$ with vertex $\sum_{r-2}$.

Suppose now $\sum_{r-1} \subset \sum_{N-1}$. Let $\sum_{N-r}$ be fixed $(N-r)$-flat skew to $\sum_{r-1}$ and let $Q_{N-r} = Q_N \cap \sum_{N-r}$. There is one-one correspondence

---

[†] This proof is given by Dowling [21].

between hyperplanes $\Sigma_{N-1}$ containing $\Sigma_{r-1}$ and (N-r-1)-flats $\Sigma_{N-r-1}$, of $\Sigma_{N-r}$, such that $\Sigma_{N-r-1} \sim \Sigma_{N-1}$ if and only if $\Sigma_{N-r-1}=\Sigma_{N-1}\cap\Sigma_{N-r}$. Since $Q_{N-r}$ is non-degenerate we can apply Theorem (3.2.5) to obtain the numbers $N_2, N_3, \ldots, N_4$ of (N-r-1)-flats $\Sigma_{N-r-1}$ in $\Sigma_{N-r}$ for which $Q_{N-r-1}=Q_{N-r}\cap\Sigma_{N-r-1}$ is a cone of order one, a non-degenerate hyperbolic quadric, and a non-degenerate elliptic quadric, respectively. But if $\Sigma_{N-1}\sim\Sigma_{N-r-1}$, and $Q_{N-r-1}$ is a cone of order s in $\Sigma_{N-r}$, then $Q_{N-1}$ is a cone of order r+s in $\Sigma_{N-1}$. The vertex $\Sigma_{r+s-1}$ of $Q_{N-1}$ is the join of $\Sigma_r$ and $\Sigma_s$, where $\Sigma_s$ is the vertex of $Q_{N-r-1}$. The proof is completed by noting that if C is the nucleus of polarity of $Q_{N-r-1}$, then $C = \Sigma_r \cap \Sigma_{N-r}$, and hence $C \in \Sigma_{N-r-1}$ if and only if $\Sigma_r \subset \Sigma_{N-1}$.

Theorem (3.3.3). Let $Q_N$ be a cone of order r in PG(N,q) where at least one of N-r and q is odd. Let $\Sigma_{r-1}$ be the vertex of $Q_N$. Let $\Sigma_{N-1}$ be any (N-1)-flat and define $Q_{N-1} = Q_N \cap \Sigma_{N-1}$. Then if $\Sigma_{r-1} \not\subset \Sigma_{N-1}$,

1. $Q_{N-1}$ is a cone of order r-1;

otherwise

2. $Q_{N-1}$ is a cone of order r or r+1.

Proof. If $\Sigma_{r-1} \not\subset \Sigma_{N-1}$ then $\Sigma_{N-1}$ contains an (N-r)-flat, $\Sigma_{N-r}$ skew to $\Sigma_{r-1}$. The quadric $Q_{N-r} = Q_N \cap \Sigma_{N-r}$ is non-degenerate in $\Sigma_{N-r}$; and $Q_{N-1}$ consists of all points on the lines joining points of $Q_{N-r}$ to the points of $\Sigma_{r-2} = \Sigma_{N-1} \cap \Sigma_{r-1}$. Hence $Q_{N-1}$ is a cone of order (r-1) with vertex $\Sigma_{r-2}$.

The number of (N-1)-flats not containing $\Sigma_{r-1}$ is

$$= \phi(N,N-1,q) - \phi(N-r,N-r-1,q)$$

$$= \frac{q^{N+1}-1}{q-1} - \frac{q^{N-r+1}-1}{q-1}$$

$$= (q^{N+1} - q^{N-r+1})/(q-1). \qquad (3.3.3)$$

Suppose now $\sum_{r-1} \subset \sum_{N-1}$. Let $\sum_{N-r}$ be a fixed $(N-r)$-flat skew to $\sum_{r-1}$ and let $Q_{N-r} = Q_N \cap \sum_{N-r}$. Then there is one-one correspondence between hyperplanes $\sum_{N-1}$ containing $\sum_{r-1}$ and $(N-r-1)$-flats $\sum_{N-r-1}$ of $\sum_{N-r}$, such that $\sum_{N-r-1} \sim \sum_{N-1}$ if and only if $\sum_{N-r-1} = \sum_{N-1} \cap \sum_{N-r}$. Since $Q_{N-r}$ is non-degenerate we can apply theorem (3.2.1) and theorem (3.2.6) to show that

$$Q_{N-r-1} = Q_{N-r} \cap \sum_{N-r-1}$$

is either a cone of order one or non-degenerate, as there is one-one correspondence between $(N-r-1)$-flats and polar spaces with respect to $Q_{N-r}$. Hence we have $|Q_{N-r}|$ $(N-r-1)$-flats for which $Q_{N-r-1}$ is a cone of order one and $\phi(N-r, N-r-1, q) - |Q_{N-r}|$ $(N-r-1)$-flats for which $Q_{N-r-1}$ is a non-degenerate cone.

But if $\sum_{N-1} \sim \sum_{N-r-1}$ and $Q_{N-r-1}$ is a cone of order $s$ in $\sum_{N-r}$ then $Q_{N-1}$ is a cone of order $r+s$ in $\sum_{N-1}$. Therefore, we have $|Q_{N-r}|$ $(N-1)$-flats that intersect $Q_N$ in a cone of order $(r+1)$; and $\phi(N-r,N-r-1,q) - |Q_{N-r}|$ $(N-1)$-flats that intersect $Q_N$ in a cone of order $r$.

Corollary 1. In theorem (3.3.3) if $N-r = 2s$ and $q$ is an odd prime power then there are $(q^{2s}-1)/(q-1)$ $(N-1)$-flat spaces that intersect $Q_N$ in a cone of order $(r+1)$; $(q^{2s}-q^s)/2$ $(N-1)$-flat spaces in an elliptic cone of order $r$; and $(q^{2s}+q^s)/2$ $(N-1)$-flats in an hyperbolic cone of order $r$.

**Proof.** From theorem (3.3.3), the number of $(N-1)$-flats that intersect $Q_N$ in a cone of order $(r+1)$ is $|Q_{N-r}|$. Since $N-r = 2s$, we have

$$|Q_{N-r}| = (q^{2s}-1)/(q-1).$$

So the number of $(N-1)$-flats that intersect $Q_N$ in a cone of order $r$ is

$$\begin{aligned} &= \phi(N-r,N-r-1,q) - |Q_{N-r}| \\ &= \frac{q^{2s+1}-1}{q-1} - \frac{q^{2s}-1}{q-1} \\ &= q^{2s} . \end{aligned}$$

But out of the $q^{2s}$ polar spaces with respect to $Q_{N-r}$, $q^s(q^s+1)/2$ polar spaces will intersect $Q_{N-r}$ in a non-degenerate hyperbolic quadric and the remaining $q^s(q^s-1)/2$ polar spaces will intersect $Q_{N-r}$ in a non-degenerate elliptic quadric[†]. Hence there are $q^s(q^s+1)/2$ polar spaces with respect to $Q_N$ that intersect $Q_N$ in a hyperbolic cone of order $r$ and $q^s(q^s-1)/2$ polar spaces that intersect $Q_N$ in a elliptic cone of order $r$.

**Corollary 2.** In theorem (3.3.3), if $N-r = 2s+1$ and $Q_N$ is an elliptic cone of order $r$, then there are $(q^{s+1}+1)(q^s-1)/(q-1)$ $(N-1)$-flat spaces that intersect $Q_N$ in a elliptic cone of order $(r+1)$; and $q^{2s+1}+q^s$ $(N-1)$-flat spaces that intersect in a cone of order $r$.

**Proof.** From theorem (3.2.8), as $N-r = 2s+1$,

$$Q_{N-r} \cap \textstyle\sum_{N-r-1} = Q_{N-r-1}$$

---

[†] Dowling [21].

is either an elliptic cone of order one or a non-degenerate quadric. Now the result follows from theorem (3.3.3).

Corollary 3. In theorem (3.3.3), if $N-r = 2s+1$ and $Q_N$ is an hyperbolic cone of order $r$, then there are $(q^{s+1}-1)(q^s+1)/(q-1)$ $(N-1)$-flat spaces that intersect $Q_N$ in an Hyperbolic cone of order $(r+1)$ and $(q^{2s+1}-q^s)$ $(N-1)$-flat spaces that intersect in a cone of order $r$.

Proof. From theorem (3.2.7), as $N-r = 2s+1$,

$$Q_{N-r} \cap \Sigma_{N-r-1} = Q_{N-r-1}$$

is either an hyperbolic cone of order one or a non-degenerate quadric. Now the result follows from theorem (3.3.3).

Using these corollaries we obtain the following lemmas. These are nothing but summaries of the preceeding results.

Lemma (3.3.4). Let $Q_{2k+1}$ be a cone of order one in $PG(2k+1,q)$, $q$ any prime power. Define $Q_{2k} = Q_{2k+1} \cap \Sigma_{2k}$, where $\Sigma_{2k}$ is a 2k-flat space. Then we can partition the 2k-flat spaces into four groups, according to the order and nature of the quadric $Q_{2k}$.

| Order of $Q_{2k}$ | number of 2k-flat spaces in the corresponding group |
|---|---|
| Non-degenerate | $q^{2k+1}$ |
| Hyperbolic cone of order one | $(q^{2k}+q^k)/2$ |
| Elliptic cone of order one | $(q^{2k}-q^k)/2$ |
| Cone of order two | $(q^{2k}-1)/(q-1)$. |

Proof. If $q$ is an odd prime power, then the result follows from theorem (3.3.3) and corollary 1. If $q$ is an even prime power, then the result follows from theorem (3.3.2).

<u>Lemma (3.3.5)</u>. Let $Q_{2k}$ be a hyperbolic cone of order one in PG(2k,q). Then we can partition the (2k-1)-flat spaces into three groups according to the nature and order of the quadric $Q_{2k-1} = Q_{2k} \cap \sum_{2k-1}$. Further

| order of $Q_{2k}$ | number of (2k-1)-flats in the corresponding group |
|---|---|
| non-degenerate | $q^{2k}$ |
| cone of order one | $q^{2k-1}-q^{k-1}$ |
| hyperbolic cone of order two | $(q^k-1)(q^{k-1}+1)/(q-1)$ |

The result follows from theorem (3.3.3) and corollary 3.

<u>Lemma (3.3.6)</u>. Let $Q_{2k}$ be an elliptic cone of order one in PG(2k,q). Then we can partition the (2k-1)-flat spaces into three groups according to the nature and order of $Q_{2k-1} = Q_{2k} \cap \sum_{2k-1}$, where $\sum_{2k-1}$ is a (2k-1)-flat space. Further

| order of $Q_{2k}$ | number of (2k-1)-flat spaces in the corresponding group |
|---|---|
| non-degenerate | $q^{2k}$ |
| cone of order one | $q^{2k-1} + q^{k-1}$ |
| elliptic cone of order two | $(q^k+1)(q^{k-1}-1)/(q-1)$. |

The result follows from theorem (3.3.3) and corollary 2.

<u>Theorem (3.3.7)</u>. Let $Q_N$ be a non-degenerate quadric in PG(N,q). Let $\sum_{N-m}$ be any (N-m)-flat. Then $Q_{N-m} = Q_N \cap \sum_{N-m}$ is a cone of order at most m, where in particular a non-degenerate quadric is considered as a cone of order zero.

**Proof.** For (N-1)-flat spaces, the result follows from theorem (3.2.1), theorem (3.2.4) and theorem (3.2.6). Suppose the result is true up to (N-m+1)-flat spaces. Let $\sum_{N-m}$ be any (N-m)-flat space. There exists at least one (N-m+1)-flat space containing $\sum_{N-m}$. Let $\sum_{N-m+1}$ be one such (N-m+1)-flat space. Define, $Q_{N-m+1}=Q_N \cap \sum_{N-m+1}$. Then by induction $Q_{N-m+1}$ is a cone of order at most (m-1). That is $Q_{N-m+1}$ is a cone of order at most (m-1) in $\sum_{N-m+1}$ and $\sum_{N-m}$ is a (N-m+1-1)-flat space in $\sum_{N-m+1}$. Hence from theorem (3.3.2) or from theorem (3.3.3) we have that $Q_{N-m}$ is a cone of order at most m.

**Lemma (3.3.8).** Let

$$Q_N: \sum_{0\le i\le j\le N} b_{ij}x_ix_j = 0$$

be a non-degenerate quadric in PG(N,q), where at least one of N and q is odd. If $\lambda$ is a non-null element of GF(q), the quadric

$$Q_{N+1}: \sum_{0\le i\le j\le N} b_{ij}x_ix_j + \lambda x_{N+1}^2 = 0$$

is non-degenerate in PG(N+1,q).

**Proof**[†]. Let $B_N$ be the matrix of $Q_N$ and $B_{N+1}$ be the matrix of $Q_{N+1}$. Then

$$B_{N+1} = \begin{bmatrix} B_N & \underline{0}^T \\ \underline{0} & \lambda \end{bmatrix}$$

where $\underline{0}$ is a row vector of (N+1) zeros.

Assume first that q is odd. Then $Q_{N+1}$ is non-degenerate if the matrix

---

[†] Dowling [21].

$$B_{N+1} + B_{N+1}^T = \begin{bmatrix} B_N + B_N^T & \underline{0}^T \\ & \\ \underline{0} & 2\lambda \end{bmatrix}$$

is non-singular. Since $Q_N$ is non-degenerate and $q$ is odd, $B_N + B_N^T$ is non-singular and $2\lambda \neq 0$. Hence $B_{N+1} + B_{N+1}^T$ is non-singular.

Next suppose that $q$ is even. Then $N$ is odd and $B_N + B_N^T$ is non-singular. But $q = 2^n$ implies $2\lambda = 0$, so that the rank of $B_{N+1} + B_{N+1}^T$ is $N+1$. Since $N+1$ is even, however, the rank of $Q_{N+1}$ is one more than the rank of $B_{N+1} + B_{N+1}^T$. Hence $Q_{N+1}$ has rank $N+2$ and therefore is non-degenerate.

Theorem (3.3.9). Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$. Let $P_0$ and $P_1$ be any two points on $Q_N$. Define

$$\Sigma_{N-2} = T(P_0) \cap T(P_1).$$

Then

i) if $P_1 \notin T(P_0)$ then $\Sigma_{N-2} \cap Q_N = Q_{N-2}$ is non-degenerate in $\Sigma_{N-2}$;

ii) if $P_1 \in T(P_0)$ then $\Sigma_{N-2} \cap Q_N = Q_{N-2}$ is a cone of order two in $\Sigma_{N-2}$.

Proof. If at least one of $N$ and $q$ is odd then $T(P_0)$ and $T(P_1)$ are well defined. Also if both are even then $T(P_0)$ and $T(P_1)$ are well defined as $P_0, P_1 \in Q_N$.

Part i): Suppose $P_1 \notin T(P_0) \Rightarrow P_0 \notin T(P_1)$.

Also, $P_1 \in T(P_1)$

and $P_0 \in T(P_0)$.

Now choose $P_2, P_3, \ldots, P_N$ as the $(N-1)$ independent points of $\sum_{N-2}$. Without loss of generality we can take the row vectors of $P_i$ as

$$\underline{P}_i = (\underbrace{0,0,\ldots,0}_{i},1,\underbrace{0,\ldots,0}_{N-i}) , \qquad i = 0,1,2,\ldots,N.$$

So,

$$T(P_0): \quad \underline{P}_0 (B+B^T)\underline{x}^T = 0$$

or

$$2b_{00}x_0 + b_{01}x_1 + \ldots + b_{0N}x_N = 0$$

Since,

$$P_0, P_2, P_3, \ldots, P_N \in T(P_0)$$

$$b_{00} = b_{02} = b_{03} = \ldots = b_{0N} = 0$$

i.e. $\quad T(P_0): \quad x_1 = 0$

Similarly, $T(P_1)$ is

$$b_{01}x_0 + 2b_{11}x_1 + \quad {}_{12}x_2 + \ldots + b_{1N}x_N = 0$$

Since

$$P_1, P_2, \ldots, P_N \in T(P_1)$$

we have $\quad T(P_1): \quad x_0 = 0$

So,

$$\sum_{N-2} = T(P_1) \cap T(P_0): \quad x_1 = 0, \quad x_0 = 0$$

Therefore

$$Q_N: \quad \underline{x}\, B\, \underline{x}^T = 0$$

or

$$(x_0, x_1, \ldots, x_N) \begin{pmatrix} 0 & b_{01} & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & b_{22} & b_{21} & \ldots & b_{2N} \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & 0 & 0 & \ldots & b_{NN} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \cdot \\ x_N \end{pmatrix} = 0$$

i.e. $b_{01}x_0x_1 + Q(x_2,x_3,\dots,x_N) = 0$

Hence

$$Q_N \cap \Sigma_{N-2}: \; x_0 = 0, \; x_1 = 0, \; Q(x_2,\dots,x_N) = 0.$$

That is $Q_{N-2} = Q_N \cap \Sigma_{N-2}$ is non-degenerate in $\Sigma_{N-2}$.

Part ii): Suppose $P_1 \in T(P_0)$

Then $P_0 \in T(P_1)$

Also, $P_1 \in T(P_1)$ and $P_0 \in T(P_0)$ as $P_0, P_1 \in Q_N$.

So, $P_0, P_1 \in \Sigma_{N-2}$.

Let $P_2$ be a point in $T(P_0) - \Sigma_{N-2}$,

and $P_3$ be a point in $T(P_1) - \Sigma_{N-2}$.

Without loss of generality we can take the row vector of $P_i$ as

$$\underline{P}_i = (\underbrace{0,0,0,\dots,0}_{i},1,\underbrace{0,\dots,0}_{N-i}), \qquad i = 0,1,2,\dots,N.$$

Hence

$$T(P_0): \; \underline{P}_0(B+B^T)\underline{x}^T = 0.$$

can be written as

$$2b_{00}x_0 + b_{01}x_1 + \dots + b_{0N}x_N = 0\,.$$

Since $P_0,P_1,P_2,P_4,\dots,P_N$ are $N$ independent points in $T(P_0)$ we have

$$b_{00} = b_{01} = b_{02} = b_{04} = b_{05} = \dots = b_{0N} = 0\,.$$

Hence $T(P_0): \; x_3 = 0$.

Similarly we can show that

$$T(P_1): \; x_2 = 0\,.$$

that is

$$b_{01} = b_{11} = b_{13} = b_{14} = \dots = b_{1N} = 0\,.$$

Hence the quadric $Q_N$ can be written as

$$x_2(b_{12}x_1 + b_{22}x_2 + b_{23}x_3 + \dots + b_{2N}x_N)$$
$$+ x_3(b_{03}x_0 + b_{33}x_3 + b_{34}x_4 + \dots + b_{3N}x_N)$$
$$+ Q(x_4, x_5, \dots, x_N) = 0$$

Hence

$$Q_{N-2} = Q_N \cap \Sigma_{N-2}$$

can be written as

$$x_2 = 0,\ x_3 = 0,\ Q(x_4, x_5, \dots, x_N) = 0.$$

That is $Q_{N-2}$ is a cone of order two. For $Q_{N-2}$ equation has only (N-1) variables and by theorem (3.3.8) $Q_{N-2}$ can be at most a cone of order two.

Theorem (3.3.10). Let $Q_N$ be a non-degenerate quadric in PG(N,q), where at least one of N and q is odd. Let $P_0$ be any point on $Q_N$ and $P_1$ be any point not on $Q_N$. Define $\Sigma_{N-2} = T(P_0) \cap T(P_1)$. Then

1. if $P_1 \notin T(P_0)$ then $\Sigma_{N-2} \cap Q_N = Q_{N-2}$ is non-degenerate in $\Sigma_{N-2}$,
2. if $P_1 \in T(P_0)$ then $\Sigma_{N-2} \cap Q_N = Q_{N-2}$ is a cone of order one.

Proof. Since at least one of N and q is odd, there is a one-one correspondence between polar spaces and (N-1)-flat spaces. Hence $T(P_0)$ and $T(P_1)$ are unique.

Part 1.

$$P_0 \notin T(P_0) \quad \text{as} \quad P_0 \in Q_N$$

Also,

$$P_1 \notin T(P_0) \Rightarrow P_0 \notin T(P_1).$$

Case a): $q \neq 2^n$, n integer.

Then

$$P_1 \notin T(P_1) \quad \text{as} \quad P_1 \notin Q_N.$$

Let $P_1^*$ be a point in $T(P_1) - \Sigma_{N-2}$. We can choose $P_2, P_3, \ldots, P_N$ $(N-1)$ points in $\Sigma_{N-2}$ such that $P_0, P_1^*, P_2, \ldots, P_N$ form a set of independent points. With out loss of generality, we can choose the vectors corresponding to these points as unit vectors, i.e.

$$\underline{P}_i = (0,0,\ldots,0,1,0,\ldots,0), \qquad i = 0,2,3,\ldots,N$$

$$\underline{P}_1^* = (0,1,0,\ldots,0).$$

$$T(P_0): \quad \underline{P}_0(B+B^T)\underline{x}^T = 0$$

i.e. $$2b_{00}x_0 + b_{01}x_1 + \ldots + b_{0N}x_N = 0$$

$$P_0, P_2, P_3, \ldots, P_N \in T(P_0)$$

$$T(P_0): \quad x_1 = 0$$

Also,

$$T(P_1): \quad \underline{P}_1(B+B^T)\underline{x}^T = 0$$

$$\Sigma_{N-2} = T(P_1) \cap T(P_0): \quad x_0 = 0, \; x_1 = 0, \quad \text{as} \quad P_0, P_1^* \in \Sigma_{N-2}.$$

$$Q_N: \quad \underline{x}\, B\, \underline{x}^T = 0$$

i.e. $$x_1(b_{01}x_0 + b_{11}x_1 + \ldots + b_{1N}x_N) + Q(x_2, x_3, \ldots, x_N) = 0.$$

It is easy to see that in $\Sigma_{N-2}$, $Q(x_2, x_3, \ldots, x_N)$ is non-degenerate.

Case b) $\quad q = 2^n$, $n$ integer.

Then $P_1 \in T(P_1)$ but $P_1 \notin Q_N$.

So, let $P_0, P_1, P_2, \ldots, P_N$ be $(N+1)$ independent points such that $P_2, P_3, \ldots, P_N$ belong to $\Sigma_{N-2}$. Without loss of generality, we can take

$$\underline{P}_i = (0,\ldots,0,1,0,\ldots,0), \qquad i = 0,1,2,\ldots,N.$$

Hence $$T(P_0): \quad x_1 = 0$$

and $$T(P_1): \quad \underline{P}_1(B+B^T)\underline{x}^T = 0$$

or

$$b_{01}x_0 + 2b_{11}x_1 + b_{12}x_2 + \ldots + b_{1N}x_N = 0$$

Since $P_1, P_2, \ldots, P_N \in T(P_1)$ we have

$$x_0 = 0$$

But $b_{11} \neq 0$ as $P_1 \notin Q_N$.

$$Q_N: \underline{x}\, B\, \underline{x}^T = 0$$

i.e. $$b_{01}x_0x_1 + b_{11}x_1^2 + Q(x_2, x_3, \ldots, x_N) = 0.$$

Therefore

$$Q_N \cap \textstyle\sum_{N-2}: \quad x_0 = 0,\ x_1 = 0,\ Q(x_2, \ldots, x_N) = 0$$

Hence the result.

<u>Part 2</u>: $P_1 \in T(P_0) \Rightarrow P_0 \in T(P_1)$.

Also, $P_0 \in T(P_0)$ as $P_0 \in Q_N$.

<u>Case a)</u>: $q \neq 2^n$. $n$ an integer

$$P_1 \notin T(P_1) \quad \text{as} \quad P_1 \notin Q_N.$$

Let $P_2$ be a point in $T(P_1) - \sum_{N-2}$. Let $P_0, P_3, P_4, \ldots, P_N$ be $(N-1)$ independent points in $\sum_{N-2}$. Without loss of generality we can take

$$\underline{P}_i = (0,0,\ldots,0,1,0,\ldots,0), \quad i = 0,1,2,\ldots,N.$$

Then

$$T(P_0): \quad 2b_{00}x_0 + b_{01}x_1 + \ldots + b_{0N}x_N = 0$$

But

$$b_{00} = b_{01} = b_{03} = b_{04} = \ldots = b_{0N} = 0 \quad \text{as} \quad P_0, P_1, P_3, P_4, \ldots, P_N \in T(P_0).$$

That is $T(P_0): \quad x_2 = 0$.

Similarly $T(P_1): \quad x_1 = 0$.

Therefore $\sum_{N-2}: \quad x_1 = 0, \quad x_2 = 0$

$$Q_N: \underline{x}\, B\, \underline{x}^T = 0$$

or

$$b_{02}x_0x_2 + b_{11}x_1^2 + x_2(b_{22}x_2+b_{23}x_3+\ldots+b_{2N}x_N) + Q(x_3,x_4,\ldots,x_N) = 0$$

That is

$$Q_N \cap \textstyle\sum_{N-2}\colon \quad x_1 = 0,\ x_2 = 0,\ Q(x_3,x_4,\ldots,x_N) = 0$$

But $Q(x_3,x_4,\ldots,x_N)$ is non-degenerate in $x_0 = 0$, $x_1 = 0$, $x_2 = 0$ (N-3)-flat space. Therefore

$$Q(x_3,x_4,\ldots,x_N) = 0, \quad x_1 = 0, \quad x_2 = 0$$

is a cone of order one with vertex $P_0$.

<u>Case b)</u> $q = 2^n$, $n$ integer

So, $P_1 \in T(P_1)$ .

Also $P_1 \in Q_N$ (Hypothesis)

That is $b_{11} \neq 0$

Also, $P_0 \in T(P_0)$ & $P_0 \in Q_N \Rightarrow b_{00} = 0$.

Let $P_2 \in T(P_0) - \sum_{N-2}$

and $P_3 \in T(P_1) - \sum_{N-2}$.

Choose points $P_4,P_5,\ldots,P_N$ in $\sum_{N-2}$ such that $P_0,P_1,P_2,P_3,P_4,\ldots,P_N$ are independent. Without loss of generality we can take

$$\underline{P_i} = (0,0,\ldots,0,1,0,\ldots,0), \qquad i = 0,1,2,\ldots,N.$$

Then $T(P_0)\colon \quad b_{01}x_1 + b_{02}x_2 +\ldots+ b_{0N}x_N = 0.$

But $P_0,P_1,P_2,P_4,P_5,\ldots,P_N \in Q_N$

So, $b_{01} = b_{02} = b_{04} = b_{05} =\ldots= b_{0N} = 0$, but $b_{03} \neq 0$.

So $T(P_0)\colon \quad x_3 = 0$

Similarly $T(P_1)\colon \quad x_2 = 0$

or

$$b_{13} = b_{14} =\ldots= b_{1N} = 0 \quad \text{but} \quad b_{12} \neq 0.$$

That is

$$\textstyle\sum_{N-2}: \quad x_2 = 0, \quad x_3 = 0.$$

$$\begin{aligned} Q_N: \quad b_{11}x_1^2 &+ x_2(b_{12}x_1+b_{22}x_2+\ldots+b_{2N}x_N) \\ &+ x_3(b_{03}x_0+b_{33}x_3+\ldots+b_{3N}x_N) \\ &+ Q(x_4,x_5,\ldots,x_N) = 0 \end{aligned}$$

Hence

$$\textstyle\sum_{N-2} \cap Q_N: \quad x_2 = 0,\ x_3 = 0,\ b_{11}x_1^2 + Q(x_4,x_5,\ldots,x_N) = 0.$$

It is easy to see that

$$x_0 = 0,\ x_1 = 0,\ x_2 = 0,\ x_3 = 0,\quad Q(x_4,\ldots,x_N) = 0$$

is non-degenerate in $\{x_0 = 0,\ x_1 = 0,\ x_2 = 0,\ x_3 = 0\}$ flat. Hence, by lemma (3.3.8),

$$x_0 = 0,\ x_2 = 0,\ x_3 = 0,\ b_{11}x_1^2 + Q(x_4,x_5,\ldots,x_N) = 0$$

is non-degenerate in $\{x_0 = 0,\ x_2 = 0,\ x_3 = 0\}$ flat as $N$ is odd. Therefore

$$x_2 = 0,\ x_3 = 0,\ b_{11}x_1^2 + Q(x_4,x_5,\ldots,x_N) = 0$$

is a cone of order one in $\{x_2 = 0,\ x_3 = 0\}$.

Theorem (3.3.11)[†]. Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$, where $q$ is odd, and let $P_0$ and $P_1$ be two points not on $Q_N$. Define $\sum_{N-2} = T(P_0) \cap T(P_1)$ and let $\sum_{N-1}$ be the $(N-1)$-flat containing $P_0$ and $\sum_{N-2}$. Then

1) if $P_1 \in \sum_{N-1}$, $\sum_{N-2} \cap Q_N$ is a cone of order one in $\sum_{N-2}$ with vertex at the point $P_2$ where the line $(P_0P_1)$ meets $Q_N$;

---

[†] Dowling [21]

2) if $P_1 \notin \sum_{N-1}$, $\sum_{N-2} \cap Q_N$ is non-degenerate in $\sum_{N-2}$.

Theorem (3.3.12)[†]. Let $Q_{2k+1}$ be a non-degenerate quadric in $PG(2k+1, q)$, where $q$ is even, and let $P_0$ and $P_1$ be two points not on $Q_{2k+1}$. Define $\sum_{2k-1} = T(P_0) \cap T(P_1)$. Then

1) if $P_1 \in T(P_0)$, $\sum_{2k-1} \cap Q_{2k+1}$ is a cone in $\sum_{2k-1}$ with vertex at the point where the line $(P_0P_1)$ meets $Q_{2k+1}$,

2) if $P_1 \notin T(P_0)$, then $\sum_{2k-1} \cap Q_{2k+1}$ is non-degenerate in $\sum_{2k-1}$.

Theorem (3.3.13). Let $Q_{2k}$ be a non-degenerate quadric in $PG(2k, 2^n)$, $n$ an integer. Let $P_0$ and $P_1$ be any two points not on $Q_{2k}$. Define $\sum_{2k-2} = T(P_0) \cap T(P_1)$. Then

1) if $P_1 \in T(P_0)$, $\sum_{2k-2} \cap Q_{2k}$ is a cone of order one in $\sum_{2k-2}$;

2) if $P_1 \notin T(P_0)$, $\sum_{2k-2} \cap Q_{2k}$ is non-degenerate in $\sum_{2k-2}$.

Proof. From lemma (3.2.2) and from lemma (3.2.3) we have that, $T(P_0)$ and $T(P_1)$ can be expressed as the polar spaces of some particular points on the quadric $Q_{2k}$. That is there exists points $P_0^*$ and $P_1^*$ on $Q_{2k}$ such that

$$T(P_0^*) \equiv T(P_0)$$

and

$$T(P_1^*) \equiv T(P_1).$$

Now, the result follows from theorem (3.3.9).

---

[†] Dowling [21]

**Lemma (3.3.14).** Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$. Let $P_0$ and $P_1$ be any two points (not on the same line) in $PG(N,q)$. Define $\Sigma_{N-2} = T(P_1) \cap T(P_0)$. If $P_1 \notin T(P_0)$, then $\Sigma_{N-2} \cap Q_N$ is non-degenerate in $\Sigma_{N-2}$.

**Proof.** The result follows from theorems (3.3.9), (3.3.10), (3.3.11), (3.3.12) and (3.3.13).

**Theorem (3.3.15).** Let $Q_{2k+1}$ be a non-degenerate hyperbolic quadric in $PG(2k+1, q)$. Let $\Sigma_{2k-1}$ be any $(2k-1)$-flat space. Define $Q_{2k-1} = \Sigma_{2k-1} \cap Q_{2k+1}$. Then we can partition the $(2k-1)$-flat spaces into four groups according to the nature and order of $Q_{2k-1}$. Further

| order of $Q_{2k-1}$ | number of $(2k-1)$-flats in the corresponding group. |
|---|---|
| hyperbolic | $q^{2k}(q^{k+1}-1)(q^k+1)/2\ (q-1)$ |
| elliptic | $q^{2k}(q^{k+1}-1)(q^k-1)/2\ (q+1)$ |
| cone of order one | $q^{k-1}(q^{2k}-1)(q^{k+1}-1)/(q-1)$ |
| hyperbolic cone of order two | $(q^{k+1}-1)(q^{k-1}+1)(q^{2k}-1)/(q-1)^2(q+1)$. |

**Proof.** For any $(2k-1)$-flat space, $\Sigma_{2k-1}$,

$$\begin{aligned} Q_{2k-1} &= Q_{2k+1} \cap \Sigma_{2k-1} \\ &= Q_{2k+1} \cap (\Sigma_{2k} \cap \Sigma_{2k-1}), \text{ where } \Sigma_{2k} \supset \Sigma_{2k-1} \\ &= (Q_{2k+1} \cap \Sigma_{2k}) \cap \Sigma_{2k-1}\ . \end{aligned}$$

Since any $(2k-1)$-flat intersects the quadric $Q_{2k+1}$ in a unique section $Q_{2k-1}$, it does not make any difference what, $\Sigma_{2k}$, $2k$-flat is considered, as long as $\Sigma_{2k}$ contains $\Sigma_{2k-1}$. Define

$$Q_{2k} = Q_{2k+1} \cap \Sigma_{2k}$$

for any 2k-flat space, $\Sigma_{2k}$. From theorem (3.2.7) it follows that we can partition the 2k-flat spaces into two groups - $C_1$: the class of 2k-flat spaces for which $Q_{2k}$ is non-degenerate; $C_2$: the class of 2k-flat spaces for which $Q_{2k}$ is a hyperbolic cone of order one. Also

$$|C_1| = (q^{2k+1}-q^k)$$

and

$$|C_2| = (q^{k+1}-1)(q^k+1)/(q-1).$$

Case 1): Let $\Sigma^o_{2k}$ be any 2k-flat in $C_1$. Define

$$Q^o_{2k} = Q_{2k+1} \cap \Sigma^o_{2k} .$$

Let $\Sigma_{2k-1}$ be any (2k-1)-flat in $\Sigma^o_{2k}$. Define

$$Q^o_{2k-1} = Q^o_{2k} \cap \Sigma_{2k-1}.$$

$Q^o_{2k}$ is non-degenerate in $\Sigma^o_{2k}$ and $\Sigma_{2k-1}$ is a (2k-1)-flat space in $\Sigma^o_{2k}$. Then we can apply theorem (3.2.5) for the case $q$ is even. Thus, the $(q^{2k+1}-1)/(q-1)$ (2k-1)-flat spaces in $\Sigma^o_{2k}$ can be partitioned into three groups according to the nature and order of $Q^o_{2k-1}$. That is

| order of $Q^o_{2k-1}$ with respect to $\Sigma^o_{2k}$ | number of (2k-1)-flats in the corresponding group |
|---|---|
| cone of order one | $(q^{2k}-1)/(q-1)$ |
| hyperbolic | $(q^{2k}+q^k)/2$ |
| elliptic | $(q^{2k}-q^k)/2$ . |

When $q$ is odd, we can apply theorem (3.2.1) and theorem (3.2.6). Then, we obtain that there are $(q^{2k}-1)/(q-1)$ (2k-1)-flat spaces in

$\sum_{2k}^{o}$ that intersect $Q_{2k}^{o}$ in a cone of order one and the remaining $q^{2k}$ (2k-1)-flat spaces in a non-degenerate quadric. Further, out of these $q^{2k}$ (2k-1)-flat spaces $q^{k}(q^{k}-1)/2$ (2k-1)-flat spaces intersect in an elliptic quadric and $q^{k}(q^{k}+1)/2$ (2k-1)-flat spaces in a hyperbolic quadric[†]. This completes the proof for case 1.

<u>Case 2)</u>: Let $\sum_{2k}^{1}$ be any 2k-flat in $C_2$. Define,

$$Q_{2k}^{1} = Q_{2k+1} \cap \sum\nolimits_{2k}^{1} .$$

Let $\sum_{2k-1}$ be any (2k-1)-flat space in $\sum_{2k}^{1}$. Then define

$$Q_{2k-1}^{1} = Q_{2k}^{1} \cap \sum\nolimits_{2k-1}.$$

$Q_{2k}^{1}$ is a hyperbolic cone of order one in $\sum_{2k}^{1}$ and $\sum_{2k-1}$ is a (2k-1)-flat space in $\sum_{2k}$. Hence by theorem (3.3.5), we can partition the $(q^{2k+1}-1)/(q-1)$ (2k-1)-flat spacesof $\sum_{2k}^{1}$ into three distinct groups according to the nature and order of $Q_{2k-1}^{1}$; i.e.

| order of $Q_{2k-1}^{1}$ | number of (2k-1)-flats in the corresponding group |
|---|---|
| hyperbolic | $q^{2k}$ |
| cone of order one | $(q^{2k-1}-q^{k-1})$ |
| hyperbolic cone of order two | $(q^{k}-1)(q^{k-1}+1)/(q-1)$ . |

Now, we shall count the number of (2k-1)-flat spaces for which $Q_{2k-1} = Q_{2k} \cap \sum_{2k-1}$ is a cone of order one. Firstly, in class $C_1$, in each 2k-flat there are $(q^{2k}-1)/(q-1)$ (2k-1)-flat spaces for which $Q_{2k-1}$ is a cone of order one. Also, in $C_2$, there are $(q^{2k-1}-q^{k-1})$ (2k-1)-flats. Hence there are

---

[†] Dowling [21]

$$|C_1| \times \frac{(q^2-1)}{(q-1)} + |C_2| \times (q^{2k-1}-q^{k-1})$$

(2k-1)-flats for which $Q_{2k-1}$ is a cone of order one. But each (2k-1)-flat occurs in exactly (q+1) 2k-flat spaces. Hence the number of (2k-1)-flat spaces for which $Q_{2k-1}$ is a cone of order one is

$$= \left\{ \frac{(q^{2k+1}-q^k)\times(q^{2k}-1)}{(q-1)} + \frac{(q^{k+1}-1)(q^k+1)\times(q^{2k-1}-q^{k-1})}{(q-1)} \right\}/(q+1)$$

$$= q^{k-1}(q^{2k}-1)(q^{k+1}-1)/(q-1).$$

Similarly we can obtain the remaining values. The number of (2k-1)-flat spaces, for which $Q_{2k-1}$ is hyperbolic, is

$$= \left\{ |C_1| \times \frac{(q^{2k}+q^k)}{2} + |C_2| \times q^{2k} \right\} / (q+1)$$

$$= \frac{1}{2(q-1)} \times q^{2k}(q^{k+1}-1)(q^k+1).$$

The number of (2k-1)-flat spaces, for which $Q_{2k-1}$ is elliptic, is

$$= |C_1| \times \frac{(q^{2k}-q^k)}{2(q+1)}$$

$$= q^{2k}(q^{k+1}-1)(q^k-1)/2(q+1).$$

Finally, the number of (2k-1)-flats, for which $Q_{2k-1}$ is hyperbolic cone of order two, is

$$= |C_2| \times \frac{(q^k-1)(q^{k-1}+1)}{(q-1)(q+1)}$$

$$= (q^{k+1}-1)(q^{k-1}+1)(q^{2k}-1)/(q-1)^2(q+1).$$

<u>Theorem (3.3.16)</u>. Let $Q_{2k+1}$ be an lliptic quadric in PG(2k+1, q) and $\sum_{2k-1}$ be any (2k-1)-flat space. Define $Q_{2k-1} = Q_{2k+1} \cap \sum_{2k-1}$. Then we can partition the (2k-1)-flat spaces into four groups according to the nature and order of $Q_{2k-1}$. Further

| order of $Q_{2k-1}$ | number of (2k-1)-flat spaces in the corresponding group. |
|---|---|
| yperbolic | $q^{2k}(q^{k+1}+1)(q^{k}+1)/2(q+1)$ |
| lliptic | $q^{2k}(q^{k+1}+1)(q^{k}-1)/2(q-1)$ |
| cone of order one | $q^{k-1}(q^{k+1}+1)(q^{2k}-1)/(q-1)$ |
| lliptic cone of order two | $(q^{k-1}-1)(q^{k+1}+1)(q^{2k}-1)/(q-1)^{2}(q+1)$. |

<u>Proof</u>. The proof of this result is exactly the same as that of theorem (3.3.15).

<u>Theorem (3.3.17)</u>. Let $Q_{2k}$ be a non-degenerate quadric in PG(2k, q). Let $\sum_{2k-2}$ be any (2k-2)-flat space. Define $Q_{2k-2} = \sum_{2k-2} \cap Q_{2k}$. Then the (2k-2)-flat spaces can be partitioned into four distinct groups according to the nature and order of the quadric $Q_{2k-2}$. Further

| order and nature of $Q_{2k-2}$ | number of (2k-2)-flat spaces in the corresponding group |
|---|---|
| non-degenerate | $q^{2k}(q^{2k}-1)/(q^{2}-1)$ |
| hyperbolic cone of order one | $q^{k-1}(q^{2k}-1)(q^{k-1}+1)/2(q-1)$ |
| elliptic cone of order one | $q^{k-1}(q^{k-1}-1)(q^{2k}-1)/2(q-1)$ |
| cone of order two | $(q^{2k}-1)(q^{2k-2}-1)/(q^{2}-1)(q-1)$ . |

<u>Proof</u>. The proof of this result is exactly the same as that of theorem (3.3.15).

Let $Q_N$ be a non-degenerate quadric in $PG(N,q)$. Let $\sum_{N-m}$ be any $(N-m)$-flat space. Define $Q_{N-m} = \sum_{N-m} \cap Q_N$. Let $\eta(N,N-m,r,q)$ be the number of $(N-m)$-flats in $PG(N,q)$ that intersect $Q_N$ in a cone of order zero. In particular, by a cone of order zero we mean a non-degenerate quadric, and by a cone of order "-1" we mean that the intersection is empty. That is $r \geq -1$. Also

$$\eta(N,N,r,q) = \begin{cases} 0 & \text{if } r \neq 0 \\ 1 & \text{if } r = 1 \end{cases}$$

From theorem (3.3.7) we have

$$\eta(N,N-m,r,q) = 0 \quad \text{if} \quad r > m.$$

Hence $\quad -1 \leq r \leq m.$

Let $\sum^o_{N-m+1}$ be any $(N-m+1)$-flat space in $PG(N,q)$. Define $Q^o_{N-m+1} = Q_N \cap \sum^o_{N-m+1}$. Let $Q^o_{N-m+1}$ be a cone of order $r$. Then for any $(N-m)$-flat space in $\sum^o_{N-m+1}$, $Q_{N-m}$ be either a cone of order $r-1$ or $r$ or $r+1$. Let $\alpha_i(N-m,\ r)$ denote the number of $(N-m)$-flats in $\sum^o_{N-m+1}$ such that $Q_{N-m} = Q^o_{N-m+1} \cap \sum_{N-m}$ is a cone of order $r-2+i$ , $\quad (i = 1,2,3)$.

## Theorem (3.3.18).

$$\eta(N,N-m,r,q) = \sum_{i=1}^{3} \eta(N,N-m+1,r-2+i,q) \times \alpha_{4-i}(N-m,r-2+i) \times \left(\frac{q-1}{q^m-1}\right)$$

Proof. Assume the result up to $m-1$. Now we shall prove the result for $m$. Let $\sum^o_{N-m+1}$ be any $(N-m+1)$-flat and $Q^o_{N-m+1} = Q_N \cap \sum^o_{N-m+1}$. Suppose $Q^o_{N-m+1}$ be a cone of order $r$ in $\sum^o_{N-m+1}$, where $-1 \leq r \leq (m-1)$. If $r = -1$, then $Q_{N-m}$ is clearly empty.

From the assumption, there are exactly $\eta(N,N-m+1,r,q)$ (N-m)-flat spaces that intersect $Q_N$ in a cone of order $r$.

Consider any (N-m)-flat $\sum_{N-m}$, in $\sum^{o}_{N-m+1}$. By theorem (2.3.2) and theorem (3.3.3) $Q_{N-m}$ is either a cone of order $r-1$ or $r$ or $r+1$. In $\sum^{o}_{N-m+1}$ there are $\alpha_i(N-m,r)$ (N-m)-flat spaces that intersect $Q_{N-m+1}$ in a cone of order $r-2+i$, $(i = 1,2,3)$. Therefore the total number of (N-m)-flat spaces that intersect $Q_N$ in a cone of order $r$,

$$\begin{aligned} T = {} & \eta(N,N-m+1,r-1,q) \times \alpha_3(N-m,r-1) \\ & + \eta(N,N-m+1,r,q) \times \alpha_2(N-m,r) \\ & + \eta(N,N-m+1,r+1,q) \times \alpha_1(N-m,r) \end{aligned}$$

If a (N-m+1)-flat intersects $Q_N$ in a cone of order greater than $r+1$ or less than $r-1$, then the (N-m)-flats, in these (N-m+1)-flat spaces, do not intersect $Q_N$ in a cone of order $r$. In $T$ each (N-m)-flat is counted $(q^m-1)/(q-1)$ times. Hence the number of (N-m)-flats in $PG(N,q)$ that intersect $Q_N$ in a cone of order $r$ is

$$\eta(N,N-m,r,q) = T \times \frac{q-1}{q^m-1}$$

The values of $\alpha_i(N-m,r)$, $i = 1,2,3$ are given by theorem (3.3.2) and theorem (3.3.3). The result for $m = 1,2$ is obtained in sections (3.2) and (3.3).

## 3.4. Retrieval schemes based on quadrics

Let $A_1, A_2, A_3, \ldots, A_\ell$ be $\ell$ attributes each with two levels. Denote the buckets by $B_1, B_2, \ldots, B_b$. Let $k_i$ be the size of i-th bucket $B_i$.

<u>Theorem (3.4.1)</u>. Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a non-degenerate quadric in PG$(2s, q=2^n)$ which is oriented toward 2s-fold queries. It involves $\ell = (q^{2s}-1)/(q-1)$ attributes and $b = (q^{2s+1}-1) \div (q-1)$ buckets.

<u>Proof</u>. Consider a non-degenerate quadric $Q_{2s}$ in PG$(2s,q=2^n)$. The quadric has exactly $(q^{2s}-1)/(q-1)$ points. Identify the points of the quadric with the attributes. Further correspond the intersections of the quadric with $(2s-1)$-flats to the buckets. Hence corresponding to each $(2s-1)$-flat we have a bucket, that is $b = (q^{2s+1}-1)/(q-1)$. Since $Q_{2s-1} = Q_{2s} \cap \sum_{2s-1}$ is not necessarily of the same order and nature for each $(2s-1)$-flat, the block sizes are unequal. From theorem (3.2.5) we have the block sizes.

Consider any 2s points in $Q_{2s}$. Then there exists a $(2s-1)$-flat, $\sum_{2s-1}$, containing the 2s points. If all the 2s points are independent then there exists exactly one $(2s-1)$-flat; otherwise there exists more than one $(2s-1)$-flats. Hence the filing scheme satisfies all queries of size 2s.

Now we shall describe a storing rule and retrieval rule for the above filing scheme. Storing rule: Each bucket is identified with $Q_{2s-1} = \sum_{2s-1} \cap Q_{2s}$, where $\sum_{2s-1}$ is $(2s-1)$-flat space. Hence there is a one-one correspondence between $(2s-1)$-flats and buckets. So, the coefficients of the equation of a $(2s-1)$-flat are taken as the identification number of the corresponding bucket.

Suppose a bucket contains k attributes $A_1, A_2, \ldots, A_k$. A record will be stored in the bucket if it has 2s attributes in common with $A_1, A_2, \ldots, A_k$ $(k \geq 2s)$. A further refinement in the storage can be made by subdividing the buckets into subbuckets such that corresponding to each 2s-fold query associated with a bucket there corresponds a subbucket. Accession numbers can be assigned to the subbuckets such that there is no duplication within a bucket. In this case more than one subbucket will have to be searched to retrieve records pertaining to a particular query.

Retrieval rule: Retrieval of records pertaining to a given 2s-fold query is performed in two stages. First the attributes are identified with the corresponding points of geometry. Than a (2s-1)-flat containing these points is obtained. Once the (2s-1)-flat is determined, the bucket can be located by matching the identification numbers. Finally, the subbucket is determined and all the accession numbers in the subbucket give the required records. Some additional subbuckets may have to be searched if a record is stored exactly once in a bucket.

Example (3.4.1). Let $s = 2$ and $q = 2$. Then $\ell = 2^4-1 = 15$. Let the attributes correspond to the points of the quadric

$$x_0x_1 + x_2x_3 + x_4^2 = 0$$

in PG(4,2). That is

$$\begin{array}{lll} A_1 = 10000, & A_6 = 10010, & A_{12} = 11011, \\ A_2 = 01000, & A_7 = 01100, & A_{12} = 00111, \\ A_3 = 00100, & A_8 = 01010, & A_{13} = 01111, \\ A_4 = 00010, & A_9 = 11001, & A_{14} = 10111, \\ A_5 = 10100, & A_{10} = 11001, & A_{15} = 11110 . \end{array} \tag{3.4.1}$$

Hence the blocks and the attributes associated with them are,

| No. | Identification | Linear equation | Attributes |
|---|---|---|---|
| $B_1$ | 10000 | $x_0 = 0$ | $A_2, A_3, A_4, A_7, A_8, A_{12}, A_{13}$ |
| $B_2$ | 01000 | $x_1 = 0$ | $A_1, A_3, A_4, A_5, A_6, A_{12}, A_{14}$ |
| $B_3$ | 00100 | $x_2 = 0$ | $A_1, A_2, A_4, A_6, A_8, A_9, A_{11}$ |
| $B_4$ | 00010 | $x_3 = 0$ | $A_1, A_2, A_3, A_5, A_7, A_9, A_{10}$ |
| $B_5$ | 10100 | $x_0+x_2 = 0$ | $A_2, A_4, A_5, A_8, A_{10}, A_{14}, A_{15}$ |
| $B_6$ | 01010 | $x_1+x_3 = 0$ | $A_1, A_3, A_5, A_8, A_{11}, A_{13}, A_{15}$ |
| $B_7$ | 11000 | $x_0+x_1 = 0$ | $A_3, A_4, A_9, A_{10}, A_{11}, A_{12}, A_{15}$ |
| $B_8$ | 10010 | $x_0+x_3 = 0$ | $A_2, A_3, A_6, A_7, A_{10}, A_{13}, A_{15}$ |
| $B_9$ | 01100 | $x_1+x_2 = 0$ | $A_1, A_4, A_6, A_7, A_{10}, A_{13}, A_{15}$ |
| $B_{10}$ | 00110 | $x_2+x_3 = 0$ | $A_1, A_2, A_9, A_{12}, A_{13}, A_{14}, A_{15}$ |
| $B_{11}$ | 11100 | $x_0+x_1+x_2 = 0$ | $A_4, A_5, A_7, A_9, A_{11}, A_{13}, A_{14}$ |
| $B_{12}$ | 11010 | $x_0+x_1+x_3 = 0$ | $A_3, A_6, A_8, A_9, A_{10}, A_{13}, A_{14}$ |
| $B_{13}$ | 10110 | $x_0+x_2+x_3 = 0$ | $A_2, A_5, A_6, A_{10}, A_{11}, A_{12}, A_{13}$ |
| $B_{14}$ | 01110 | $x_1+x_2+x_3 = 0$ | $A_1, A_7, A_8, A_{10}, A_{11}, A_{12}, A_{14}$ |
| $B_{15}$ | 11110 | $x_0+x_1+x_2+x_3 = 0$ | $A_5, A_6, A_7, A_8, A_9, A_{12}, A_{15}$ |
| $B_{16}$ | 00001 | $x_4 = 0$ | $A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_{15}$ |
| $B_{17}$ | 10001 | $x_0+x_4 = 0$ | $A_2, A_3, A_4, A_7, A_8, A_9, A_{10}, A_{11}, A_{14}$ |
| $B_{18}$ | 01001 | $x_1+x_4 = 0$ | $A_1, A_3, A_4, A_5, A_6, A_9, A_{10}, A_{11}, A_{13}$ |
| $B_{19}$ | 00101 | $x_2+x_4 = 0$ | $A_2, A_3, A_4, A_7, A_8, A_9, A_{10}, A_{11}, A_{14}$ |
| $B_{20}$ | 00011 | $x_3+x_4 = 0$ | $A_1, A_2, A_3, A_5, A_7, A_{11}, A_{12}, A_{13}, A_{14}$ |
| $B_{21}$ | 10101 | $x_0+x_2+x_4 = 0$ | $A_2, A_4, A_5, A_8, A_9, A_{11}, A_{12}, A_{13}, A_{15}$ |
| $B_{22}$ | 10011 | $x_0+x_3+x_4 = 0$ | $A_2, A_3, A_6, A_7, A_9, A_{10}, A_{12}, A_{13}, A_{15}$ |
| $B_{23}$ | 01101 | $x_1+x_2+x_4 = 0$ | $A_1, A_4, A_6, A_7, A_9, A_{11}, A_{12}, A_{14}, A_{15}$ |
| $B_{24}$ | 01011 | $x_1+x_3+x_4 = 0$ | $A_1, A_3, A_5, A_8, A_9, A_{10}, A_{12}, A_{14}, A_{15}$ |
| $B_{25}$ | 11111 | $x_0+x_1+x_2+x_4 = 0$ | $A_5, A_6, A_7, A_8, A_{10}, A_{11}, A_{13}, A_{14}, A_{15}$ |
| $B_{26}$ | 11001 | $x_0+x_1+x_4 = 0$ | $A_3, A_4, A_{13}, A_{14}, A_{15}$ |
| $B_{27}$ | 00111 | $x_2+x_3+x_4 = 0$ | $A_1, A_2, A_{10}, A_{11}, A_{15}$ |
| $B_{28}$ | 11101 | $x_0+x_1+x_2+x_4 = 0$ | $A_4, A_5, A_7, A_{10}, A_{12}$ |
| $B_{29}$ | 11011 | $x_0+x_1+x_3+x_4 = 0$ | $A_3, A_6, A_8, A_{11}, A_{12}$ |
| $B_{30}$ | 10111 | $x_0+x_2+x_3+x_4 = 0$ | $A_3, A_5, A_6, A_9, A_{14}$ |
| $B_{31}$ | 01111 | $x_1+x_2+x_3+x_4 = 0$ | $A_1, A_7, A_8, A_9, A_{13}$ . |

There are 31 blocks - 15 buckets are of size 7, 10 of size 9 and 6 of size 15. The identification numbers of the buckets are the coefficients of the general equation

$$c_0x_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = 0$$

where $c_i \in GF(2)$. The subbuckets in a bucket are formed by considering all possible 4-plets in that bucket. The identification of a subbucket is formed by concatenation of binary representation of the 4 attributes corresponding to it. For example,

| No. | Bucket Id. | Subbucket | Accession number of records |
|---|---|---|---|
| $B_{26}$ | 11001 | $A_3A_4A_{13}A_{14}$ | .... .... .. .... |
| | | $A_3A_4A_{13}A_{15}$ | .... .... .. .... |
| | | $A_3A_4A_{14}A_{15}$ | .... .... .. .... |
| | | $A_3A_{13}A_{14}A_{15}$ | .... .... .. .... |
| | | $A_4A_{13}A_{14}A_{15}$ | .... .... .. .... |
| ............ | ............ | ............ | ............ |
| $B_{27}$ | 00111 | $A_1A_2A_{10}A_{11}$ | .... .... .. .... |
| | | $A_1A_2A_{10}A_{15}$ | .... .... .. .... |
| | | $A_1A_2A_{11}A_{15}$ | .... .... .. .... |
| | | $A_1A_{10}A_{11}A_{15}$ | .... .... .. .... |
| | | $A_2A_{10}A_{11}A_{15}$ | .... .... .. .... |
| ............ | ............ | ............ | ............ |

And so on. A record is stored in a bucket if it contains at least one point of the bucket. For example a record with

$$f(I) = (\overline{A}_1, A_2, A_3, A_4, \overline{A}_5, \overline{A}_6, \overline{A}_7, \overline{A}_8, \overline{A}_9, \overline{A}_{10}, \overline{A}_{11}, \overline{A}_{12}, \overline{A}_{13}, A_{14}, A_{15})$$

(where $A_i$ indicates the presence of i-th attribute and $\overline{A}_i$ indicates the absence of i-th attribute) is stored in the bucket 11001 i.e. $B_{26}$, since

$$|B_{26} \cap f(I)| \geq 4.$$

This record is also stored in the buckets 10001($B_{17}$), 00001($B_{16}$), 10010($B_8$), and 10100($B_5$). Further, if the record has more than four points in common with a bucket, then it will be stored in more than one subbucket. This duplication, in a bucket, is avoided by ordering the subbuckets. The record is stored in the first subbucket that comes across. But a four tuple may occur in more than one bucket. That is the same subbucket will occur in more than one bucket. This is avoided by using the chaining procedure between subbuckets. For example, the buckets $B_{16}$ and $B_{17}$, have five points in common i.e. $A_2, A_3, A_4$, $A_7$ and $A_8$. Hence the subbuckets

$$A_2A_3A_4A_7$$
$$A_2A_3A_4A_8$$
$$A_2A_3A_7A_8$$
$$A_2A_4A_7A_8$$
$$A_3A_4A_7A_8$$

occur in both these buckets. So instead of storing the accession numbers of the records which have these attributes, the addresses of these subbuckets in $B_{16}$ are stored in $B_{17}$. This avoides a lot of repetitions of the accession numbers.

Given any query of order four the attributes of this query are converted into the points of the geometry. Then a 3-flat containing these points is determined. The rest of procedure consists of identifying the buckets and subbuckets. As an example, suppose the query is $\{A_1A_3A_4A_5\}$. The corresponding points, from (3.4.1), are

$A_1$ = 10000 $A_4$ = 00010
$A_3$ = 00100 $A_5$ = 10100

These points are not independent. Hence there is more than one 3-flat containing these four points. Suppose the equation of the 3-flat is

$$c_0x_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 = 0 \qquad (3.4.2)$$

Then,

$$c_0 = 0, \quad c_2 = 0, \quad c_3 = 0$$

Hence,

$$c_1x_1 + x_4x_4 = 0 .$$

The 3-flats containing the points are

$$x_1 = 0; \quad x_4 = 0; \quad \text{and} \quad x_1+x_4 = 0 .$$

Without loss of generality we can consider the 3-flat, $x_1+x_4 = 0$. Hence the bucket containing these points is 01001 i.e. $B_{18}$. The subbucket is obtained by matching $A_1A_3A_4A_5$ with subbucket identification numbers. As the records are stored in the first subbucket that comes across, we have to search all the preceeding subbuckets in the bucket to obtain all the records satisfying the query. But in $B_{18}$, $\{A_1A_3A_4A_5\}$ subbucket is the first subbucket. Hence we do not have to search any other subbuckets to obtain all the accession numbers satisfying the query.

Theorem (3.4.2). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a hyperbolic quadric in PG(2s+1, q) which is oriented toward (2s+1)-fold queries. It involves $\ell = \frac{q^{2s+1}-1}{q-1} + q^s$ attributes, $b = (q^{2s+2}-1) \div (q-1)$ buckets such that $(q^{2s+1}-q^s)$ buckets have size $(q^{2s}-1)/(q-1)$ and the remaining $\frac{(q^{s+1}-1)(q^s+1)}{(q-1)}$ buckets have size $(\frac{q^{2s}-1}{q-1}) + q^s$.

Proof. The proof of this result is exactly the same as that of the previous result.

Example (3.4.2). Let $s = 1$ and $q = 2$. Consider the quadric,

$Q_3$: $x_0x_1 + x_2x_3 = 0$.

Then, $\ell = 9$, $b = 15$.

$A_1 = 1000$, $A_4 = 1010$, $A_7 = 0101$,

$A_2 = 0100$, $A_5 = 1001$, $A_8 = 0001$,

$A_3 = 0010$, $A_6 = 0110$, $A_9 = 1111$ .

The buckets and the corresponding attributes are:

| Bucket no. | Identification | Attributes | Subbuckets |
|---|---|---|---|
| $B_1$ | 1000 | $A_2, A_3, A_6, A_7, A_8$ | all triplets |
| $B_2$ | 0100 | $A_1, A_3, A_4, A_5, A_8$ | all triplets |
| $B_3$ | 0010 | $A_1, A_2, A_5, A_7, A_8$ | " |
| $B_4$ | 0001 | $A_1, A_2, A_3, A_4, A_6$ | " |
| $B_5$ | 1010 | $A_2, A_4, A_7, A_8, A_9$ | " |
| $B_6$ | 1001 | $A_2, A_3, A_5, A_6, A_9$ | " |
| $B_7$ | 0110 | $A_1, A_5, A_6, A_8, A_9$ | " |
| $B_8$ | 0101 | $A_1, A_3, A_4, A_7, A_9$ | " |
| $B_9$ | 1111 | $A_4, A_5, A_6, A_7, A_9$ | " |
| $B_{10}$ | 1100 | $A_3, A_8, A_9$ | $A_3A_8A_9$ |
| $B_{11}$ | 0011 | $A_1, A_2, A_9$ | $A_1A_2A_9$ |
| $B_{12}$ | 1110 | $A_4, A_6, A_8$ | $A_4A_6A_8$ |
| $B_{13}$ | 1101 | $A_3, A_5, A_7$ | $A_3A_5A_7$ |
| $B_{14}$ | 1011 | $A_2, A_4, A_5$ | $A_2A_4A_5$ |
| $B_{15}$ | 0111 | $A_1, A_6, A_7$ | $A_1A_6A_7$ |

Each of the first nine buckets have each ten subbuckets. So there are 96 subbuckets. But there are only $\binom{9}{3} = 84$ triplets. The storing and retrieval rules are exactly the same as in example (3.4.2).

Theorem (3.4.3). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a non-degenerate elliptic quadric in PG(2s+1, q), oriented toward (2s+1)-fold queries. It involves $\ell = \frac{q^{2s+1}-1}{q-1} - q^s$ attributes, $b = (\frac{q^{2s+2}-1}{q-1})$ buckets such that $(q^{2s+1} + q^s)$ buckets have size $(q^{2s}-1)/(q-1)$ and the remaining have size $((q^{2s}-1)/(q-1)) - q^s$.

When $q = 2$ and $s = 1$, there are five buckets which have size one. These buckets are redundant, when only the queries of size three are considered. So the number of buckets will be less than the value specified by the theorem. This will not arise for higher values of q and s.

Example (3.4.3). $q = 2$, $s = 1$.

Let $Q_3$: $x_0^2 + x_0x_1 + x_1^2 + x_2x_3 = 0$.

It has $\ell = 5$ and $b \leq 15$.

$A_1 = 0001$, $A_3 = 0111$, $A_5 = 1111$,

$A_2 = 0010$, $A_4 = 1011$.

The buckets and their corresponding attributes are:

| Bucket no. | Identification | Attributes |
|---|---|---|
| $B_1$ | 1000 | $A_1A_2A_3$ |
| $B_2$ | 0100 | $A_1A_2A_4$ |
| $B_3$ | 1100 | $A_1A_2A_5$ |
| $B_4$ | 0011 | $A_3A_4A_5$ |
| $B_5$ | 1010 | $A_1A_4A_5$ |
| $B_6$ | 1001 | $A_2A_4A_5$ |
| $B_7$ | 0110 | $A_1A_3A_5$ |
| $B_8$ | 0101 | $A_2A_3A_5$ |
| $B_9$ | 1110 | $A_1A_3A_4$ |
| $B_{10}$ | 1101 | $A_2A_3A_4$ |

........................................

| Bucket no. | Identification | Attributes |
|---|---|---|
| | 0010 | $A_1$ |
| | 0001 | $A_2$ |
| | 0011 | $A_3$ |
| | 0111 | $A_4$ |
| | 1111 | $A_5$ . |

In fact we can delete all those buckets whose size is less than query size. The above filing scheme is optimum. In the sense, each triplet occurs exactly once.

Theorem (3.4.4). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on non-degenerate hyperbolic quadric in PG(2s+1, q) which is oriented toward 2s-fold queries. It involves $\{(q^{s+1}-1)/(q-1)\} + q^s$ attributes, and $b = (q^{2s+2}-1)(q^{2s+1}-1)/(q^2-1)(q-1)$ buckets such that $q^{2s}(q^{s+1}-1) \times (q^s+1)/2(q-1)$ buckets have size $\{(q^{2s-1}-1)/(q-1)\} - q^{s-1}$; $q^{2s}(q^{s+1}-1)(q^s-1)/2(q+1)$ have size $\{(q^{2s-1}-1)/(q-1)\}-q^{s-1}$; $q^{s-1}(q^{2s}-1)(q^{s+1}-1) \div (q-1)$ have size $(q^{2s-1}-1)/(q-1)$ and finally the remaining buckets have size $(q^{2s-1} + q^{s+1} - q^s-1)/(q-1)$.

Proof. The attributes are identified with the points of a non-degenerate hyperbolic quadric in PG(2s+1, q). The intersections of the quadric with (2s-1)-flat spaces are considered as buckets. So, the number of buckets is equal to the number of (2s-1)-flat spaces in PG(2s+1, q). The actual sizes of these buckets are given by theorem (3.3.15). The storage and retrieval rules can be taken as before.

Example (3.4.4). Let $s = 1$ and $q = 2$. Then $\ell = 9$, $b \le 35$.

$$Q_3: \quad x_0x_1 + x_2x_3 = 0 \qquad \text{(as in example 3.4.2)} .$$

We can neglect 11 buckets, as their sizes are less than or equal to one.

Hence there will be only 24 buckets. These satisfy queries of size two. The buckets and their corresponding attributes are:

| Bucket no. | Identification | Attributes | Subbuckets |
|---|---|---|---|
| $B_1$ | 1000 0100 | $A_3, A_8$ | |
| $B_2$ | 1000 0110 | $A_6, A_8$ | |
| $B_3$ | 1000 0101 | $A_3, A_7$ | |
| $B_4$ | 1000 0111 | $A_6, A_7$ | |
| $B_5$ | 0100 1010 | $A_4, A_8$ | |
| $B_6$ | 0100 1001 | $A_3, A_5$ | |
| $B_7$ | 0100 1011 | $A_4, A_5$ | |
| $B_8$ | 0010 0001 | $A_1, A_2$ | |
| $B_9$ | 0010 1001 | $A_2, A_5$ | |
| $B_{10}$ | 0010 0101 | $A_1, A_4$ | |
| $B_{11}$ | 0010 1101 | $A_5, A_7$ | |
| $B_{12}$ | 0001 1010 | $A_2, A_4$ | |
| $B_{13}$ | 0001 0110 | $A_1, A_6$ | |
| $B_{14}$ | 0001 1110 | $A_4, A_6$ | |
| $B_{15}$ | 1100 1010 | $A_8, A_9$ | |
| $B_{16}$ | 1100 1001 | $A_3, A_9$ | |
| $B_{17}$ | 0110 0101 | $A_1, A_9$ | |
| $B_{18}$ | 1010 1001 | $A_2, A_9$ | |
| $B_{19}$ | 1000 0010 | $A_2, A_7, A_8$ | $A_2A_7; A_2A_8; A_7A_8$ |
| $B_{20}$ | 1000 0001 | $A_2, A_3, A_6$ | $A_2A_3; A_2A_6; A_3A_6$ |
| $B_{21}$ | 0100 0010 | $A_1, A_5, A_8$ | $A_1A_5; A_1A_8; A_5A_8$ |
| $B_{22}$ | 0100 0001 | $A_1, A_3, A_4$ | $A_1A_3; A_1A_4; A_3A_4$ |
| $B_{23}$ | 0110 1001 | $A_5, A_6, A_9$ | $A_5A_6; A_5A_9; A_6A_9$ |
| $B_{24}$ | 1010 0101 | $A_4, A_7, A_9$ | $A_4A_9; A_4A_7; A_7A_9$ |

The total number of subbuckets is 36 and the total number of queries is 36. Hence corresponding to each query there is exactly one subbucket. The storage scheme can be further simplified by forming super-buckets. For example, we can group all the buckets with "1000" as the first four symbols of its identification in one super-bucket. Also "1000" can be taken as its identification. So, we have

| Super-bucket | bucket | Subbucket Id. | Record Accession nos. |
|---|---|---|---|
| 1000 | 0100 | $A_3A_8$ | |
| | 0110 | $A_6A_8$ | |
| | 0101 | $A_3A_7$ | |
| | 0111 | $A_6A_7$ | |
| | 0010 | $A_2A_7$; $A_2A_8$; $A_7A_8$ | |
| | 0001 | $A_2A_3$; $A_2A_6$; $A_3A_6$ | |
| 0100 | 1010 | $A_4A_8$ | |
| | 1001 | $A_3A_5$ | |
| | 1011 | $A_4A_5$ | |
| | 0010 | $A_1A_5$; $A_1A_8$; $A_5A_8$ | |
| | 0001 | $A_1A_3$; $A_1A_4$; $A_3A_4$ | |
| 0010 | 0001 | $A_1A_2$ | |
| | 1001 | $A_2A_5$ | |
| | 0101 | $A_1A_4$ | |
| | 1101 | $A_5A_7$ | |
| 0001 | 1010 | $A_2A_4$ | |
| | 0110 | $A_1A_6$ | |
| | 1110 | $A_4A_6$ | |
| 1100 | 1010 | $A_8A_9$ | |
| | 1001 | $A_3A_9$ | |
| 1010 | 1001 | $A_2A_9$ | |
| | 0101 | $A_4A_9$; $A_4A_7$; $A_7A_9$ | |
| 0110 | 0101 | $A_1A_9$ | |
| | 1001 | $A_5A_6$; $A_5A_9$; $A_6A_9$ | |

Theorem (3.4.5). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a non-degenerate elliptic quadric in PG(2s+1, q) which is oriented toward 2s-fold queries. It involves $\ell = \frac{q^{2s+1}-1}{q-1} - q^s$ attributes, $b = (q^{2s+2}-1)(q^{2s+1}-1)/(q-1)^2(q+1)$ buckets.

Proof. The proof of this result is similar to that of theorem (3.4.4). The bucket sizes are given by theorem (3.3.16).

Theorem (3.4.6). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a non-degenerate quadric in PG(2s, q) which is oriented toward (2s-1)-fold queries. It involves $\ell = (q^{2s}-1)/(q-1)$ attributes, and $b = \frac{(q^{2s+1}-1)(q^{2s}-1)}{(q^2-1)(q-1)}$ buckets.

The bucket sizes are obtained from theorem (3.3.17).

Theorem (3.4.7). Given that the retrieval pertains to only one level of each attribute, there exists a filing scheme based on a non-degenerate quadric in PG(2s, q) which is oriented toward 2s-fold queries. It involves $\ell = q^{2s}$ attributes and $b = (q^{2s+1}-1)/(q-1)$ buckets. Of these b buckets, $(q^{2s}-1)/(q-1)$ have size $q^{2s}-1$; $(q^{2s}+q^s)/2$ have size $(q^{2s-1}-q^{s-1})$; and $(q^{2s}-q^s)/2$ have size $(q^{2s-1}+q^{s-1})$.

Proof. Consider a non-degenerate quadric $Q_{2s}$ in PG(2s, q). Let $\sum_{2s}$ denote the whole space, PG(2s, q). Define,

$$R_{2s} = \sum_{2s} - Q_{2s} .$$

$R_{2s}$ denotes the set of points not on the quadric $Q_{2s}$. The attributes are identified with the points on $R_{2s}$. $R_{2s}$ has exactly $q^{2s}$ points.

The buckets are identified with the intersections of $R_{2s}$ with (2s-1)-flat spaces. Define

$$\begin{aligned} R_{2s-1} &= R_{2s} \cap \textstyle\sum_{2s-1} \\ &= \textstyle\sum_{2s-1} - (Q_{2s} \cap \textstyle\sum_{2s-1}) \end{aligned}$$

The properties of $Q_{2s} \cap \sum_{2s-1}$ are given by theorem (3.2.5). Hence $R_{2s-1}$ is known. Hence the result.

Example (3.4.5). $s = 2$ and $q = 2$

Then $\ell = 2^4 = 16$.

$$Q_4: \ x_0x_1 + x_2x_3 + x_4^2$$

The points and the corresponding attributes are

$A_1 = 11000$,
$A_2 = 11010$,
$A_3 = 11100$,
$A_4 = 11001$,
$A_5 = 01001$,
$A_6 = 10101$,
$A_7 = 01101$,
$A_8 = 00001$,
$A_9 = 00110$,
$A_{10} = 01110$,
$A_{11} = 10110$,
$A_{12} = 00011$,
$A_{13} = 00101$,
$A_{14} = 10011$,
$A_{15} = 01011$,
$A_{16} = 11111$ .

There are 31 blocks - 15 have size 8, 10 have size 6 and the remaining 6 blocks have size 10. The same design can be obtained from EG(4,2). It has exactly 30 blocks each of size 8. The design obtained will be more useful, if some 4-tuples are retrieved more frequently than others, than the design obtained from EG(4,2).

## 3.5 Multiple level attributes

Consider a non-degenerate quadric $Q_N$ in PG(N,q). Let $\pi_1, \pi_2, \ldots, \pi_{q+1}$ be (q+1) (N-1)-flat spaces containing the same (N-2)-flat space, $\sum_{N-2}^{o}$. Define,

$$Q^i_{N-1} = Q_N \cap \pi_i \,, \qquad i = 1,2,\ldots,q+1 \tag{3.5.1}$$

$$Q^o_{N-2} = Q_N \cap \sum^o_{N-2} \tag{3.5.2}$$

also,

$$R_N = Q_N - Q^o_{N-2} \tag{3.5.3}$$

and

$$R^i_{N-1} = Q^i_{N-1} - Q^o_{N-2}, \quad i = 1,2,\ldots,q+1 \tag{3.5.4}$$

or,

$$R^i_{N-1} = \pi_i \cap R_N \tag{3.5.5}$$

or

$$R^i_{N-1} = (Q_N \cap \pi_i) - (Q_N \cap \sum^o_{N-2}) \tag{3.5.6}$$

From the results of section (3.3) and section (3.4) the cardinalities of $R^i_{N-1}$ $(i = 1,2,\ldots,q+1)$ are known. Let

$$n_i = |R^i_{N-1}|. \tag{3.5.7}$$

The $n_i$'s are not necessarily equal.

Let $\sum_{t-1}$ be any (t-1)-flat space in PG(N,q). Define

$$R_{t-1} = R_N \cap \sum_{t-1} \tag{3.5.8}$$

Consider any $t$ points $P_0,P_1,\ldots,P_{t-1}$ from $R_N$ such that no two points are contained in the same (N-1)-flat, $\pi_i$ $(i = 1,2,\ldots,(q+1))$. There exists at least one (t-1)-flat containing these $t$ points. Let $\sum_{t-1}$ be one such (t-1)-flat space. Then there exists a $R_{t-1}$ which contains these $t$ points, where

$$R_{t-1} = \sum_{t-1} \cap R_N.$$

Hence, given any $t$ points, there exists a $R_{t-1}$ containing them. That is, there exists a filing scheme for $(q+1)$ attributes with levels $n_1, n_2, \ldots, n_{q+1}$ such that it satisfies queries of size $t$. The buckets of this filing scheme are the intersections of $R_N$ with $(t-1)$-flat spaces.

Example (3.5.1). Let $q = 2$, $N = 3$, & $t = 2$.

$$Q_3: \quad x_0x_1 + x_2x_3 = 0$$

It has exactly 9 points;

$$\begin{array}{ccc} 1000 & 0100 & 0010 \\ 0001 & 0110 & 0101 \\ 1010 & 1001 & 1111 \ . \end{array}$$

Let

$$\pi_1: \ x_0 + x_1 = 0$$

$$\pi_2: \ x_2 + x_3 = 0$$

and $$\pi_3: \ x_0 + x_1 + x_2 + x_3 = 0 \ .$$

Further

$$\pi_1 \cap Q_3: \ 0010, 0001, 1111$$

$$\pi_2 \cap Q_3: \ 1000, 0100, 1111$$

$$\pi_3 \cap Q_3: \ 1010, 1001, 0110, 0101, 1111 \ .$$

So

$$R_2^1: \ 0010, 0001$$

$$R_2^2: \ 1000, 0100$$

and $$R_2^3: \ 1010, 1001, 0110, 0101$$

So the levels of the attributes can be taken as

$$n_1 = 2, \ A_{11} = 0010; \ A_{12} = 0001$$

$$n_2 = 2, \ A_{21} = 1000; \ A_{22} = 0100$$

$$n_3 = 4, \ A_{31} = 1010; \ A_{32} = 1001; \ A_{33} = 0110; \ A_{34} = 0101.$$

The buckets are formed by considering lines that intersect in at least two $R_2^i$'s. There are ten such lines. For storage convenience the buckets are grouped into super-buckets. i.e.

| Super-buckets (id) | buckets (id) | subbuckets | accession nos. |
|---|---|---|---|
| 1000 | 0001 | $A_{11}A_{22}$ | |
| | | $A_{11}A_{33}$ | |
| | | $A_{22}A_{33}$ | |
| | 0010 | $A_{12}A_{22}$ | |
| | | $A_{12}A_{34}$ | |
| | | $A_{22}A_{34}$ | |
| | 0101 | $A_{11}A_{34}$ | |
| | 0110 | $A_{12}A_{33}$ | |
| 0100 | 0001 | $A_{11}A_{21}$ | |
| | | $A_{11}A_{31}$ | |
| | | $A_{21}A_{31}$ | |
| | 0010 | $A_{12}A_{21}$ | |
| | | $A_{12}A_{32}$ | |
| | | $A_{21}A_{32}$ | |
| 0100 | 1010 | $A_{12}A_{31}$ | |
| | 1001 | $A_{11}A_{32}$ | |
| 0010 | 1001 | $A_{22}A_{32}$ | |
| | 0101 | $A_{21}A_{34}$ | |
| 0001 | 1010 | $A_{22}A_{31}$ | |
| | 0110 | $A_{21}A_{33}$ | |

There are exactly five super-buckets and at most four buckets in any super-bucket. This arrangement greatly reduces the retrieval time (in terms of matching).

A record is stored in a bucket exactly once. This requires chaining between subbuckets. The record is stored in the first subbucket that comes across in a bucket.

Suppose we are interested interested in retrieving all records containing the levels $A_{21}$ and $A_{32}$. The corresponding points of these levels in PG(3,2) are $A_{21} = 1000$, $A_{32} = 1001$. The line satisfying these points is

$$c_0x_0 + c_1x_1 + c_2x_2 + c_3x_3 = 0 ,$$
$$d_0x_0 + d_1x_1 + d_2x_2 + c_3x_3 = 0$$

or

$$c_0 = d_0 = 0$$
$$c_3 = d_3 = 0 .$$

So the equations of the line are

$$x_1 = 0, \quad x_2 = 0.$$

The super-bucket identification number is 0100; and that of the bucket is 0010.

The retrieval time for any query A, involves three components,

$$\tau(A) = T_1 + T_2 + T_3,$$

where $T_1$ is the time required to solve the linear equations; $T_2$ is the time required to locate the super-bucket, bucket and the subbucket by matching the identification numbers; and $T_3$ is the time required to retrieve the actual records. $T_1$ and $T_3$ are dependent, mainly, on the computer system. Let $\tau_1$ and $\tau_3$ be the maximum possible values of $T_1$ and $T_3$ for any query. $T_2$ depends on the distribution of records. Suppose the records are uniformly distributed. Then the average number of comparisons required is

$$(1\times1\times1 + 1\times1\times2 +\ldots+ 4\times2\times1)/20 = 4.8.$$

So the average retrieval time

$$\tau \leq \tau_1 + \tau_3 + 4.8\epsilon ,$$

where $\epsilon$ is the machine time required for each single comparison. When we use simple inverted filing scheme, we need on the average 10.5 comparisons.

Example (3.5.2). $N = 4$, $q = 2$, $\ell = 3$

$$Q_4: \quad x_0x_1 + x_2x_3 + x_4^2 = 0$$

It has exactly 15 points. The three 3-flat spaces can be taken as

$$\pi_1: \quad x_0 = 0 \quad ; \quad \pi_2: \quad x_1 = 0; \quad \pi_3: \quad x_0+x_1 = 0 \; .$$

Further,

$R_3^1$: 01000; 01100; 01010; 01111

$R_3^2$: 10000; 10100; 10010; 10111

$R_3^3$: 11001; 11101; 11011; 11110

So, we have

$n_1 = 4$; $A_{11} = 01000$ $A_{12} = 01100$; $A_{13} = 01010$; $A_{14} = 01111$

$n_2 = 4$; $A_{21} = 10000$ $A_{22} = 10100$; $A_{23} = 10010$; $A_{24} = 10111$

$n_3 = 4$; $A_{31} = 11001$ $A_{32} = 11101$; $A_{33} = 11011$; $A_{34} = 11110$ .

The buckets can be identified with 1-flat spaces, which intersect at least two $R_3^i$'s. The storage and retrieval rule can be as that of the previous example (3.5.1).

# CHAPTER IV

# A GENERALIZED FILING SCHEME FOR MULTIPLE LEVEL ATTRIBUTES

In this chapter we obtain filing schemes for multiple valued attributes, using linear representation of the attributes.

## 4.1 Linear representation of attributes

Let $A_1, A_2, A_3, \ldots, A_\ell$ be $\ell$ attributes with $n_1, n_2, n_3, \ldots, n_\ell$ levels respectively. We shall assume, throughout this chapter, that

$$n_i = q^{r_i}, \qquad i = 1, 2, \ldots, \ell,$$

where $q$ is a prime power and $r_i$ some positive integer. Since $n_i = q^{r_i}$ for the i-th attribute $A_i$, we can represent the $n_i$ levels by $r_i$-tuples over $GF(q)$; i.e.

$$A_{ij} = \underline{V}^i_j = (V^i_{j1}, V^i_{j2}, \ldots, V^i_{jr_i}), \quad V^i_{jm} \in GF(q)$$

where $A_{ij}$ is the j-th level of i-th attribute and $\underline{V}^i_j$ is the q - ary representation of the number $j$. The above representation is unambiguous and unique. As an example, when $q = 2$, $\ell = 2$, $n_1 = 4$, $n_2 = 8$, we have $r_1 = 2$ and $r_2 = 3$. Further

$$\begin{aligned} &A_{10} = (00), \; A_{11} = (01), \; A_{12} = (10), \; A_{13} = (11), \\ &A_{20} = (000), \; A_{21} = (001), \; A_{22} = (010), \; A_{23} = (011), \\ &A_{24} = (100), \; A_{25} = (101), \; A_{26} = (110), \; A_{27} = (111). \end{aligned}$$

From now on, in this chapter we use, mainly the q-ary representation of the levels of the attributes. Let

$$N = \sum_{i=1}^{\ell} r_i$$

and

$$p = \max(r_{i_1} + r_{i_2} + \ldots + r_{i_t})$$

where the maximum is taken over all possible t-tuples, $(i_1,\ldots,i_t)$. There are exactly $\binom{\ell}{t}$ such t-tuples.

Let

$$\underset{N \times m}{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ \vdots & \vdots & & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{Nm} \end{bmatrix} \qquad (4.1.1)$$

be a $(N\times m)$ matrix with elements of $GF(q)$. Also, the rank of $H$ is $m$. Let $r_{j.} = \sum_{k=0}^{j-1} r_k$ where $r_0 = 0$.

$$\underset{r_j \times m}{H_j} = \begin{bmatrix} h_{r_{j.}+1,1} & h_{r_{j.}+1,2} & \cdots & h_{r_{j.}+1,m} \\ h_{r_{j.}+2,1} & h_{r_{j.}+2,2} & \cdots & h_{r_{j.}+2,m} \\ \vdots & \vdots & & \vdots \\ h_{r_{j+1.},1} & h_{r_{j+1.},2} & \cdots & h_{r_{j+1.},m} \end{bmatrix} \qquad (4.1.2)$$

for $j = 1,2,\ldots,\ell$.

That is, $H_1,H_2,\ldots,H_\ell$ are submatrices of $H$. So we can write,

$$H = \begin{bmatrix} H_1 \\ \cdots\cdots \\ H_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_\ell \end{bmatrix} \qquad (4.1.3)$$

Definition (4.1.1). The matrix $H$ is said to have the property $R_t(r_1, r_2, \ldots, r_\ell)$ if for any $t$ of the submatrices $H_{i_1}, H_{i_2}, \ldots, H_{i_t}$ of $H$ the matrix

$$\underset{\left(\sum_{j=1}^{t} r_{i_j}\right) \times m}{G} = \begin{bmatrix} H_{i_1} \\ \cdots\cdots \\ H_{i_2} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_{i_t} \end{bmatrix} \qquad (4.1.4)$$

has rank $\sum_{j=1}^{t} r_{i_j}$ .

If $H$ has the property $R_t(r_1, \ldots, r_\ell)$ then $m \geq p$. We shall use the notation $R_t(\ell \times n)$ for $R_t(r_1, \ldots, r_\ell)$ when $r_1 = r_2 = \ldots = r_\ell = n$. If $r_1 = r_2 = \ldots = r_\ell = 1$, then the property $R_t(\ell \times 1)$ is called the $P_t$ property. The matrices with property $P_t$ were first considered by Bose (1947) in connection with the problem of confounding symmetrical factorial designs. Later these matrices were used in constructing error correcting codes. (Bose and Ray-Chaudhuri [10] & [11] and Peterson [26].) In later parts of this chapter, a method for obtaining matrices with property $R_t(r_1,r_2,\dots,r_\ell)$ will be described.

The i-th attribute $A_i$, $(i = 1,2,\dots,\ell)$ will be represented by a set of linear forms

$$\underset{(r_i\times m)(m\times 1)}{H_i\ \underline{x}^T} = \underset{(r_i\times 1)}{\underline{L}^{i^T}} \tag{4.1.5}$$

or in full

$$\begin{matrix} h_{r_{i\cdot}+1,1}\ x_1 + h_{r_{i\cdot}+1,2}\ x_2 + \dots + h_{r_{i\cdot}+1,m}\ x_m = L^i_1 \\ h_{r_{i\cdot}+2,1}\ x_1 + h_{r_{i\cdot}+2,2}\ x_2 + \dots + h_{r_{i\cdot}+2,m}\ x_m = L^i_2 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ h_{r_{i+1\cdot},1}\ x_1 + h_{r_{i+1\cdot},2}\ x_2 + \dots + h_{r_{i+1\cdot},m}\ x_m = L^i_{r_i} \end{matrix} \tag{4.1.6}$$

where $H_i$ is the i-th submatrix of $H$ in (4.1.3) with property $R_t(r_1,r_2,\dots,r_\ell)$.

To any query,

$$Q\begin{pmatrix} A_{i_1} & A_{i_2} & \dots & A_{i_g} \\ \underline{V}_{j_1}^{i_1} & \underline{V}_{j_2}^{i_2} & \dots & \underline{V}_{j_g}^{i_g} \end{pmatrix} \tag{4.1.7}$$

where $g \le t$, we shall associate the set of equations,

$$\underline{L}^{i_\omega} = \underline{V}_{j_\omega}^{i_\omega} \quad , \quad \omega = 1,2,\dots,g.$$

That is

$$\begin{aligned}
h_{r_{i_\omega}\cdot+1,1}\, x_1 + h_{r_{i_\omega}\cdot+1,2}\, x_2 + \cdots + h_{r_{i_\omega}\cdot+1,m}\, x_m &= V^{i_\omega}_{j_{\omega,1}} \\
h_{r_{i_\omega}\cdot+2,1}\, x_1 + h_{r_{i_\omega}\cdot+2,2}\, x_2 + \cdots + h_{r_{i_\omega}\cdot+2,m}\, x_m &= V^{i_\omega}_{j_{\omega,2}} \\
\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \quad &\qquad \vdots \\
h_{r_{i_\omega+1}\cdot,1}\, x_1 + h_{r_{i_\omega+1}\cdot,2}\, x_2 + \cdots + h_{r_{i_\omega+1}\cdot,m}\, x_m &= V^{i_\omega}_{j_\omega, r_{i_\omega}} \; .
\end{aligned} \tag{4.1.8}$$

for $\omega = 1,2,\ldots,g.$

or

$$\begin{aligned}
H^{i_1}\, \underline{x}^T &= \underline{V}^{i_1}_{j_1} \\
H^{i_2}\, \underline{x}^T &= \underline{V}^{i_1}_{j_2} \\
\cdots \quad &\cdots \; \cdots \\
H^{i_g}\, \underline{x}^T &= \underline{V}^{i_g}_{j_g}
\end{aligned} \tag{4.1.9}$$

Let

$$G = \begin{bmatrix} H^{i_1} \\ \cdots\cdots \\ H^{i_2} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H^{i_g} \end{bmatrix} \qquad \underline{V} = \begin{bmatrix} \underline{V}^{i_1}_{j_1} \\ \cdots\cdots \\ \underline{V}^{i_2}_{j_2} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \underline{V}^{i_g}_{j_g} \end{bmatrix} \tag{4.1.10}$$

So, we can rewrite (4.1.9) as

$$G\, \underline{x}^T = \underline{V} \tag{4.1.11}$$

where $G$ is $\left(\sum_{j=1}^{g} r_{i_j}\right) \times m$ matrix and $\underline{V}$ is $\left(\sum_{j=1}^{g} r_{i_j}\right) \times 1$ column.

$$\text{rank } G = \sum_{j=1}^{g} r_{i_j} = r \quad \text{(say)} \tag{4.1.12}$$

as H has property $R_t(r_1,\ldots,r_\ell)$ and $g \le t$. There exists at least one $(r\times r)$ submatrix in G with rank r. Let $\alpha_1,\alpha_2,\ldots,\alpha_r$ be the columns in this submatrix. In the equation (4.1.8) we can take the values of all the variables $x_\alpha$, $\alpha \neq \alpha_1,\alpha_2,\ldots,\alpha_r$ equal to zero and solve the resulting equations for $x_{\alpha_1},x_{\alpha_2},\ldots,x_{\alpha_r}$. Let

$$x_{\alpha_i} = u_{\alpha_i}, \qquad i = 1,2,\ldots,r.$$

We then get a solution

$$\underline{u} = (u_1,u_2,\ldots,u_m)$$

of (4.1.8), in which all the coordinates other than $\alpha_1$-th, $\alpha_2$-th, $\ldots$, $\alpha_r$-th are equal to zero.

Let $\beta_0$ be the set of m-vectors over GF(q) for which not more than p of the coordinates are non-null. The number of vectors in $\beta_0$ is

$$b_0 = 1 + (q-1)\binom{m}{1} + (q-1)^2\binom{m}{2} + \ldots + (q-1)^p\binom{m}{p} \qquad (4.1.13)$$

We have shown that there exists at least one solution for (4.1.8) in $\beta_0$, as $r \le p$. By using the "Echelon method - canonical solution" we can find an unique solution in $\beta_0$. So, without loss of generality, we take $\underline{u}$ to be the unique solution for (4.1.8). For any m-vector $\underline{u}$, the linear forms $\underline{L}^i$, given by (4.1.6) may be said to attain the value

$$\underline{L}^i(\underline{u}) = H^i\,\underline{u}^T \qquad (4.1.14)$$

at $\underline{u}$. Further,

$$\underline{L}^{i_\omega}(u) = \underline{V}_{j_\omega}^{i_\omega}, \qquad \omega = 1,2,\ldots,g.$$

So, we have,

<u>Theorem (4.1.1)</u>. To any query,

$$Q\begin{pmatrix} A_{i_1} & A_{i_2} & \cdots & A_{i_\omega} \\ \underline{V}_{j_1}^{i_1} & \underline{V}_{j_2}^{i_2} & \cdots & \underline{V}_{j_\omega}^{i_\omega} \end{pmatrix}$$

we can make correspond an unique m-vector $\underline{u}$ of $\beta_0$, namely the canonical solution of the equations associated to the query, such that the linear forms $\underline{L}^{i_1}, \underline{L}^{i_2}, \ldots, \underline{L}^{i_g}$ attain the values $\underline{V}_{j_1}^{i_1}, \underline{V}_{j_2}^{i_2}, \ldots, V_{j_g}^{i_g}$.

The number of queries of size less than or equal to $t$, is

$$\sum_{k=1}^{t} \sum_{(i_1,i_2,\ldots,i_k)} q^{r_{i_1}+r_{i_2}+\ldots+r_{i_k}} \tag{4.1.15}$$

where $\sum_{(i_1 \ldots i_k)}$ denotes summation over all possible k-tuples.

In general, the same vector $\underline{u}$ may correspond to more than one query, though it is possible that there are vectors in $\beta_0$ not corresponding to any query. Let $\beta$ be the subset of $\beta_0$, such that to any vector of $\Omega$ there corresponds at least one query. Let $b = |\beta|$.

### Example (4.1.1)

$$v = 2;\ t = 2;\ \ell = 5;\ n_i = 2,\ i = 1,2,3,4,5.$$

So, $r_i = 2$ for all $i$.

$$\underset{10\times 4}{H} = \begin{bmatrix} 0001 \\ 0010 \\ 0100 \\ 1000 \\ 0110 \\ 1011 \\ 0101 \\ 1010 \\ 0111 \\ 1110 \end{bmatrix} \tag{4.1.16}$$

$$H_1 = \begin{bmatrix} 0001 \\ 0010 \end{bmatrix}; \quad H_2 = \begin{bmatrix} 0100 \\ 1000 \end{bmatrix}; \quad H_3 = \begin{bmatrix} 0110 \\ 1011 \end{bmatrix}$$

$$H_4 = \begin{bmatrix} 0101 \\ 1010 \end{bmatrix}; \; H_5 = \begin{bmatrix} 0111 \\ 1110 \end{bmatrix}$$

The (10×4) matrix H has the property $R_2(5\times 2)$, but it does not have the property $P_4$ - i.e. any four rows of H are not independent. The linear forms and their corresponding attributes are,

| attribute set | linear form |
|---|---|
| $L_1^1$ | $x_4$ |
| $L_2^1$ | $x_3$ |
| $L_1^2$ | $x_2$ |
| $L_2^2$ | $x_1$ |
| $L_1^3$ | $x_2 + x_3$ |
| $L_2^3$ | $x_1 + x_3 + x_4$ |
| $L_1^4$ | $x_2 + x_4$ |
| $L_2^4$ | $x_1 + x_3$ |
| $L_1^5$ | $x_2 + x_3 + x_4$ |
| $L_2^5$ | $x_1 + x_2 + x_3$ |

(4.1.17)

1). Consider the query of size two,

$$Q \begin{pmatrix} A_4 & A_5 \\ \underline{V}_2^4 & \underline{V}_3^5 \end{pmatrix}$$

The associated equations are,

$$\begin{aligned} x_2 + x_4 &= V^4_{21} \\ x_1 + x_3 &= V^4_{22} \\ x_2 + x_3 + x_4 &= V^5_{31} \\ x_1 + x_2 + x_3 &= V^5_{32} \end{aligned} \tag{4.1.18}$$

Since H has rank 4, we have a unique solution

$$\begin{aligned} x_1 &= V^4_{21} + V^4_{22} + V^5_{31} \\ x_2 &= V^4_{22} + V^5_{32} \\ x_3 &= V^4_{21} + V^5_{31} \\ x_4 &= V^4_{21} + V^4_{22} + V^5_{32} \end{aligned} \tag{4.1.19}$$

But $\underline{V}^4_2 = (01)$ and $\underline{V}^5_3 = (10)$, so the actual solution for (4.1.17) is

$$\underline{u} = (0,1,1,1).$$

ii) Consider the query

$$Q \begin{pmatrix} A_3 \\ \underline{V}^3_4 \end{pmatrix}$$

Then the corresponding equations are

$$\begin{aligned} x_2 + x_3 &= V^3_{41} \\ x_1 + x_3 + x_4 &= V^3_{42} \end{aligned}$$

So

$$\begin{aligned} x_1 + x_3 + x_4 &= V^3_{42} \\ x_2 + x_3 &= V^3_{41} \end{aligned}$$

Put

$$x_3 = x_4 = 0$$

Then the solution vector is

$$\underline{u} = (V^3_{42}, V^3_{41}, 0, 0)$$

or

$$\underline{u} = (1,1,0,0) \quad \text{as} \quad \underline{V}^3_4 = (1,1).$$

The value of the linear forms for the solution vector (1,1,0,0) in (4.1.17) is

$$\underline{L}\ (\underline{u}) = (0,0,1,1,1,1,1,1,1,0).$$

## 4.2 Generalized multiple valued filing scheme

Let $\underline{u}$ be any solution vector in $\beta$. To each $\underline{u}$ there correspond an N-vector

$$B(\underline{u}) = (\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^{\ell})$$

where $\underline{b}^i$ is a $(1\times r_i)$ vector, $(i = 1,2,\ldots,\ell)$ over GF(q) such that

$$\underline{b}^i = L^i(\underline{u}) \ , \qquad i = 1,2,\ldots,\ell.$$

$B(\underline{u})$ is called the block corresponding to $\underline{u}$. Corresponding to each query there exists a vector $\underline{u}$, and hence there exists a block, $B(\underline{u})$. We shall show that the correspondence between the block $B(\underline{u})$ and the vector $\underline{u}$ is unique. Let $\underline{u}_1$ and $\underline{u}_2$ be any two distinct m-vectors in $\beta$ such that

$$B(\underline{u}_1) = B(\underline{u}_2) \ ,$$

i.e.

$$L^i_j(\underline{u}_1) = L^i_j(\underline{u}_2) \ , \qquad \text{for all } i \text{ and } j \cdot$$

Hence

$$L_j^i(\underline{u}_1-\underline{u}_2) = 0, \qquad \text{for all } i \text{ and } j. \tag{4.2.1}$$

Since the rank of $H$ is $m$, we can choose $m$ rows of $H$ such that they are independent. Without loss of generality, let

$$L_1^{i_1}, L_2^{i_1}, \ldots, L_{j_1}^{i_1}, \ldots, L_1^{i_k}, \ldots, L_{j_k}^{i_k}$$

where $\sum_{i=1}^{k} j_i = m$, be a set of $m$ independent rows in $H$. Then

$$\begin{aligned} L_1^{i_1}(\underline{u}_1-\underline{u}_2) &= 0 \\ L_2^{i_1}(\underline{u}_1-\underline{u}_2) &= 0 \\ \vdots \qquad \vdots \quad & \quad \vdots \\ L_{j_k}^{i_k}(\underline{u}_1-\underline{u}_2) &= 0 \end{aligned} \tag{4.2.2}$$

Since (4.2.2) is a set of $m$ independent homogeneous linear equations in $m$ variables, we must have

$$\underline{u}_1-\underline{u}_2 = 0$$

or

$$\underline{u}_1 = \underline{u}_2 \quad .$$

This is a contradiction, since $\underline{u}_1$ and $\underline{u}_2$ are two distinct vectors. Hence, the correspondence between m-vectors of $\beta$ and the blocks is one-one. We can now obtain a correspondence between queries and blocks. If, to the query $Q$ there corresponds an m-vector $\underline{u}$ of $\beta$, then we say that $B(\underline{u})$ is the bucket corresponding to $Q$. Hence corresponding to any query $Q$ of size less than or equal to $t$ there exists a block, but to each block there corresponds, in general, many queries.

**Theorem (4.2.1).** If the block $(\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^{\ell})$ corresponds to the query

$$Q\begin{pmatrix} A_{i_1} & A_{i_2} & \cdots & A_{i_g} \\ \underline{V}_{j_1}^{i_1} & \underline{V}_{j_2}^{i_2} & \cdots & \underline{V}_{j_g}^{i_g} \end{pmatrix} \qquad (4.2.3)$$

then

$$\begin{aligned} \underline{b}^{i_1} &= \underline{V}_{j_1}^{i_1} \\ \underline{b}^{i_2} &= \underline{V}_{j_2}^{i_2} \\ &\cdots\cdots\cdots\cdots \\ \underline{b}^{i_g} &= \underline{V}_{j_g}^{i_g} \quad . \end{aligned} \qquad (4.2.4)$$

**Proof.** From theorem (4.1.1) there exists a unique solution $\underline{u}$ an m-vector, corresponding to Q. Also,

$$\begin{aligned} L^{i_1}(\underline{u}) &= \underline{V}_{j_1}^{i_1} \\ L^{i_2}(\underline{u}) &= \underline{V}_{j_2}^{i_2} \\ &\cdots\cdots\cdots\cdots \\ L^{i_g}(\underline{u}) &= \underline{V}_{j_g}^{i_g} \end{aligned} \qquad (4.2.5)$$

Since,

$$\underline{b}^{i_k} = L^{i_k}(\underline{u}), \qquad k = 1,2,\ldots,g \ .$$

we have

$$\underline{b}^{i_k} = \underline{V}_{j_k}^{i_k} \ , \qquad k = 1,2,\ldots,g \ .$$

In the example (4.1.1) when the query is

$$Q\begin{pmatrix} A_3 \\ \underline{V}_4^3 \end{pmatrix}, \quad \text{where} \quad \underline{V}_4^3 = (1,1)$$

then

$$B(\underline{u}) = (0,0,1,1,1,1,1,1,1,0).$$

Let $M$ be the set of addresses reserved for storing the accession numbers of records. $M$ can be partitioned into $b$ subsets, corresponding to the block $B(\underline{u})$ being denoted by $M(\underline{u})$. These subsets of $M$ are called buckets. The vector $\underline{u}$ will be called the identification of the bucket $M(\underline{u})$. Consider the block $B(\underline{u})$ corresponding to the bucket $M(\underline{u})$.

$$B(\underline{u}) = (\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^{\ell})$$

where $\underline{b}^i$ is a $r_i$-vector. In fact $\underline{b}^i$ represents some level of the i-th attribute. $M(\underline{u})$ is subdivided into $2^{\ell}-1$ subsets, corresponding to the $2^{\ell}-1$ nonempty subsets of $\{\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^{\ell}\}$. These subsets of $M(\underline{u})$ are called subbuckets. The subbucket of $M(\underline{u})$ which corresponds to the subset $A^*$ of $\{\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^{\ell}\}$ may be denoted by $M(\underline{u}, A^*)$. Let $f(I) = (V^1_{j_1}, V^2_{j_2}, \ldots, V^{\ell}_{j_{\ell}})$ be the attribute vector of I-th record. Define

$$B(\underline{u}) \cap f(I) = A \qquad (4.2.5)$$

where

$$A = \{\underline{b}^i : \underline{b}^i = \underline{V}^i_{j_i}\} .$$

That is, $\underline{b}^i \in A$, if and only if the level of i-th attribute for I-th record is the same as the value attained by the linear forms $L^i$ at $\underline{u}$.

Our storage rule can now be stated as follows. The accession number of the I-th record is stored in the bucket $M(\underline{u})$ if $M(\underline{u}) \cap f(I)$ is non-empty. If $M(\underline{u}) \cap f(I) = A^*$ then the accession number of the I-th record will be stored in the subbucket $M(\underline{u}, A^*)$ of $M(\underline{u})$. By this rule a record will be stored exactly once in a bucket.

We have noted earlier that the redundancy, $\zeta$ in general depends on the distribution of the attribute records. We shall calculate the value of $\zeta$ under the hypothesis of uniform distribution, and corresponding to the storage rule described in the previous paragraph.

There are $q^N$ possible distinct N-vectors over GF(q). Each is a possible attribute vector. Suppose the number of records having any given attribute vector is $\delta$, and is independent of the vector. Then the total number of records is $\delta q^N$. If $f(I)$ is the attribute vector of the I-th record, then the accession number of this record does not appear in the bucket $M(\underline{u})$ if and only if $B(\underline{u}) \cap f(I)$ is empty. $B(\underline{u})$ being given there are

$$(q^{r_1}-1)(q^{r_2}-1) \times \ldots \times (q^{r_\ell}-1)$$

ways of choosing $f(I)$ such that $B(\underline{u}) \cap f(I) = \emptyset$. Hence there are

$$\delta \times \prod_{j=1}^{\ell} (q^{r_j}-1)$$

records whose accession numbers do not occur in the bucket $M(\underline{u})$. The number of records whose accession numbers occur in the bucket is, therefore

$$\delta q^N - \delta \prod_{j=1}^{\ell} (q^{r_j}-1) .$$

Since there are $b = |\beta|$ buckets the total number of the accession numbers stored in the storage is

$$b \times \delta \times [q^N - \prod_{j=1}^{\ell} (q^{r_j}-1)]$$

Hence,

$$\zeta = \frac{b \times \delta \times [q^N - \prod_{j=1}^{\ell} (q^{r_j}-1)]}{\delta \times q^N}$$

$$= b[1 - \prod_{j=1}^{\ell} (1 - \frac{1}{q^{r_j}} ) \qquad (4.2.6)$$

as

$$N = \sum_{j=1}^{\ell} r_j .$$

Example (4.2.1.). Continuing the example (4.1.1) consider the block, $B(\underline{u}) = (0,0,1,1,1,1,1,1,1,0)$. We can rewrite as, $B(\underline{u}) = (0,3,3,3,2)$. Also, $B(\underline{u}) = (A_{10},A_{23},A_{33},A_{43},A_{52})$. So the bucket $M(\underline{u})$ corresponding to the block $B(\underline{u})$ has the identification number $\underline{u} = (1,1,0,0)$. The subbuckets of the bucket are

$A_{10};A_{23};A_{33};A_{43};A_{52}$

$A_{10}A_{23};A_{10}A_{33};A_{10}A_{43};A_{10}A_{52}$

$A_{23}A_{33};A_{23}A_{43};A_{23}A_{52}$

$A_{33}A_{43};A_{33}A_{52}$

$A_{43}A_{52};$

$A_{10}A_{23}A_{33};A_{10}A_{23}A_{43};A_{10}A_{23}A_{52}$

$A_{10}A_{33}A_{43};A_{10}A_{33}A_{52};A_{10}A_{43}A_{52}$

$A_{23}A_{33}A_{43};A_{23}A_{33}A_{52};A_{23}A_{43}A_{52}$

$A_{33}A_{43}A_{52};$

$A_{10}A_{23}A_{33}A_{43};A_{10}A_{23}A_{33}A_{52};A_{10}A_{33}A_{43}A_{52}$

$A_{10}A_{23}A_{43}A_{52};A_{23}A_{33}A_{43}A_{52}$

Let the attribute vector corresponding to the I-th vector be

$$f(I) = (0,3,1,2,2) = (A_{10},A_{23},A_{31},A_{42},A_{52})$$

So, $$B(\underline{u}) \cap f(I) = (A_{10}, A_{23}, A_{52})$$

Hence, the accession number of the I-th record is stored in the subbucket $(A_{10}A_{23}A_{52})$.

Let

$$Q(A) = \begin{pmatrix} A_{i_1} & A_{i_2} & \cdots & A_{i_g} \\ \underline{V}_{j_1}^{i_1} & \underline{V}_{j_2}^{i_2} & \cdots & \underline{V}_{j_g}^{i_g} \end{pmatrix}, \quad g \le t$$

be any query of size less than or equal to $t$. Let $\underline{u}$ be the unique solution of (4.1.8) corresponding to the query $Q(A)$, and let $B(\underline{u}) = (\underline{b}^1, \underline{b}^2, \ldots, \underline{b}^\ell)$ be the corresponding block. Then the bucket $M(\underline{u})$ corresponding to $B(\underline{u})$ is defined to be the bucket corresponding to the query. From theorem (4.2.1)

$$\underline{b}^{i_\omega} = \underline{V}_{j_\omega}^{i_\omega}, \qquad \omega = 1, 2, \ldots, g .$$

If the I-th record satisfies the query, then the $i_1$-th, $i_2$-th, ..., $i_g$-th components of the attribute vector $f(I)$ of $I$ are

$$\underline{V}_{j_1}^{i_1}, \underline{V}_{j_2}^{i_2}, \ldots, \underline{V}_{j_g}^{i_g} .$$

That is,

$$B(\underline{u}) \cap f(I) \supset Q(A) . \qquad (4.2.7)$$

Hence, there exists a subset $A^*$ in $M(\underline{u})$ that contains the accession number of I-th record where

$$B(\underline{u}) \cap f(I) = A^* .$$

Conversely let $A^*$ be any subset such that $A^* \supset Q(A)$, the query. It is clear from our storage rule that all accession numbers of records that are stored in the subbucket $M(\underline{u}, A^*)$ also satisfy the query $Q(A)$. Hence to find all accession numbers of records satisfying the query we have to search all the subbuckets $M(\underline{u}, A^*)$, where $A^* \supset Q(A)$. That is

$$M(Q) = \cup\, M(\underline{u}, A^*) \qquad (4.2.8)$$

where the union is taken over all subsets $A^*$ that contain the query set $Q(A)$. There are exactly $2^{\ell-g}$ subsets $A^*$ that contain $Q(A)$.

So, the retrieval rule may be stated as follows: let $Q$ be any query relating to the attributes of the subset $Q(A)$ of $A$. Determine the bucket $B(\underline{u})$ corresponding to $Q$. Then the accession numbers of all records satisfying $Q$ are found in $M(Q)$ given by (4.2.8). Conversely any record whose accession number is in $M(Q)$ satisfies the query $Q$. Once the accession numbers have been found we can retrieve the complete records.

Example (4.2.2). Continuing the example (4.1.1)

Consider the query

$$Q\begin{pmatrix} A_1 & A_2 \\ \underline{V}_0^1 & \underline{V}_3^2 \end{pmatrix}$$

From (4.1.17) we have

$$\underline{u} = (1,1,0,0)$$

So,

$$B(\underline{u}) = (0,0,1,1,1,1,1,1,1,0)$$

or

$$B(\underline{u}) = (0,3,3,3,2).$$

So there are $2^{5-2} = 8$ subsets containing $(A_{10}, A_{23})$ in $B(\underline{u})$. These eight subbuckets are located by searching within the bucket $M(\underline{u})$.

The time required for retrieving the records satisfying a given query will be made up of a number of components:

$T_1$: time required for coding physical attributes into linear forms and the values of the attributes into vectors over GF(q).

$T_2$: time required for reducing the query equations into Echelon form and determining the Canonical solution $\underline{u}$, which gives the bucket identification number.

$T_3$: time needed for identifying the bucket $M(\underline{u})$.

$T_4$: time required for computing the labels of the subbuckets entering in $M(Q)$.

$T_5$: time required for identifying the subbuckets.

$T_6$: time required for actual retrieval of the records.

Thus the total time required for a query $Q(A)$ is

$$\tau(A) = T_1+T_2+T_3+T_4+T_5+T_6.$$

The components $T_1, T_2, T_4$ and $T_6$ will depend on the parameters of the computer system used. By using an ordering among bucket labels and among subbuckets within a bucket, the components $T_3$ and $T_5$ can be reduced.

The filing scheme for the case $n_1=n_2=\ldots=n_\ell$ was obtained by Bose [7]. In his method the levels are not represented by vectors.

We shall now describe a method of obtaining $(N\times m)$ matrices with elements from GF(q), which have rank $m$ and property $R_t(r_1,r_2,\ldots,r_\ell)$. Before describing the method we shall study briefly some properties of spreads.

## 4.3 Correspondence between Galois fields, vector spaces and projective spaces

Let $q$ be a prime power. Let $(N+1) = \theta(m+1)$ for some positive integers $m$, $N$, $\theta(>1)$. Consider the Galois field $GF(q^{N+1})$. It has exactly $q^{N+1}$ elements. Let $\alpha$ be a primitive root of $GF(q^{N+1})$. Then

$$\alpha^{q^{N+1}-1} = 1. \tag{4.3.1}$$

All the elements of $GF(q^{N+1})$ can be expressed in the form

$$b_0+b_1\alpha^1+b_2\alpha^2+\ldots+b_N\alpha^N$$

where $b_i \in GF(q)$ for all $i$. Similarly any element of $GF(q^{m+1})$ can be written as

$$c_0+c_1\eta^1+c_2\eta^2+\ldots+c_m\eta^m$$

where $c_i \in GF(q)$ for all $i$ and

$$\eta = \alpha^{(q^{N+1}-1)/(q^{m+1}-1)} .$$

$\eta$ is a primitive element of $GF(q^{m+1})$. Since the representation of elements of a Galois field with respect to a primitive element is unique, the correspondence

$$b_0+b_1\alpha+\ldots+b_N\alpha^N \longleftrightarrow (b_0,b_1,\ldots,b_N) \tag{4.3.2}$$

between the elements of $GF(q^{N+1})$ and $(N+1)$ - vectors over $GF(q)$ is one-one. Similarly we can obtain one-one correspondence between the elements of $GF(q^{m+1})$ and $(m+1)$-vectors over $GF(q)$:

$$c_0+c_1\eta^1+c_2\eta^2+\ldots+c_m\eta^m \longleftrightarrow (c_0,c_1,\ldots,c_m) . \tag{4.3.3}$$

The Galois field $GF(q^{N+1})$ can be viewed as an extension of $GF(q^{m+1})$, as $(N+1) = \theta(m+1)$. So the element of $GF(q^{N+1})$ can be uniquely expressed as

$$a_0+a_1\delta^1+a_2\delta^2+\ldots+a_{\theta-1}\delta^{\theta-1} \tag{4.3.4}$$

where $a_0,a_1,\ldots,a_{\theta-1}$ belong to $GF(q^{m+1})$ and $\delta$ is a primitive element of $GF(q^{N+1})$. Now the correspondence,

$$(a_0+a_1\delta^1+a_2\delta^2+\ldots+a_{\theta-1}\delta^{\theta-1}) \longleftrightarrow (a_0,a_1,\ldots,a_{\theta-1}) \qquad (4.3.5)$$

between the elements of $GF(q^{N+1})$ and $\theta$-vectors over $GF(q^{m+1})$ is one-one. By (4.3.3) each $a_i$ can be expressed as an $(m+1)$-vector over $GF(q)$. So we can correspond $\theta$-vectors of (4.3.5) and $(N+1)$-vectors of (4.3.2). That is

$$(a_0,a_1,\ldots,a_{\theta-1}) \longrightarrow (\underline{a}_0,\underline{a}_1,\ldots,\underline{a}_{\theta-1}) \qquad (4.3.6)$$

where $\underline{a}_i$ is the $(m+1)$-vector representation of $a_i$ over $GF(q)$. Conversely an $(N+1)$-vector can be made to correspond to a $\theta$-vector as follows;

$$(b_0,b_1,\ldots,b_N) \longrightarrow (a_0,a_1,\ldots,a_{\theta-1}) \qquad (4.3.7)$$

where

$$\begin{aligned} \underline{a}_0 &= (b_0,b_1,\ldots,b_m) \\ \underline{a}_1 &= (b_{m+1},b_{m+2},\ldots,b_{2m+1}) \\ &\vdots \\ \underline{a}_{\theta-1} &= (b_{(\theta-1)(m+1)},\ b_{(\theta-1)(m+1)+1},\ldots,b_N)\ . \end{aligned}$$

That is, there is a one-one correspondence between the $\theta$-vectors over $GF(q^{m+1})$ and the N-vectors over $GF(q)$.

Let $V_{N+1}$ be the $(N+1)$-dimensional vector space over $GF(q)$. Let $V_{N+1} = (b_0,b_1,\ldots,b_N)$ be any non-zero $(N+1)$-vector in $V_{N+1}$. Then $V_{N+1}$ represents a point P in $PG(N,q)$. Also the $(N+1)$-vectors $(c\ b_0,c\ b_1,\ldots,c\ b_N)$ where $c$ is any non-zero element in $GF(q)$, represent the same point P. Hence there is a one-one correspondence between the points of $PG(N,q)$ and the one dimensional subspaces of $V_{N+1}$. A basis of the one dimensional subspace is taken as a representative of the point P as a subspace is completely specified by a basis.

The null vector $(0,0,\ldots,0)$ may be regarded as corresponding to the $(-1)$-flat (the empty space) of $PG(N,q)$, where the empty space is a subspace of every k-flat of $PG(N,q)$, $k = -1,0,1,\ldots,N$. There is a one-one correspondence between k-flats of $PG(N,q)$ and $(k+1)$-dimensional subspaces of $V_{N+1}$.

Lemma (4.3.1). There is a one-one correspondence between k-flat spaces of $PG(N,q)$ and $(k+1)$-dimensional subspaces of $V_{N+1}$, the $(N+1)$-dimensional space over $GF(q)$, where $0 \leq k \leq N$.

Let $\underline{s} = (s_1,s_2,\ldots,s_\theta)$ be any $\theta$-vector in $\theta$-dimensional vector space $V_\theta$ over $GF(q^{m+1})$. Let

$$W_\theta(\underline{s}) = \{c\,\underline{s}: \quad c \in GF(q^{m+1})\}. \tag{4.3.8}$$

Then,

Lemma (4.3.2)[†]. The $(N+1)$-vectors of $V_{N+1}$ over $GF(q)$ corresponding to $q^{m+1}$ $\theta$-vectors of $W_\theta(\underline{s})$ are the elements of an $(m+1)$-dimensional subspace of $V_{N+1}$. The null element of the subspace corresponds to the null element of $V_{N+1}$.

Proof. Let $u_0,u_1,\ldots,u_{m+1}$ be any $(m+2)$ non-zero elements of $GF(q^{m+1})$. Since each $u_i$ can be uniquely expressed as an $(m+1)$-vector over $GF(q)$ there exist $c_0,c_1,\ldots,c_{m+1}$ in $GF(q)$, not all zero, such that

$$c_0u_0+c_1u_1+\ldots+c_{m+1}u_{m+1} = 0. \tag{4.3.9}$$

Hence

$$c_0u_0\underline{s}+c_1u_1\underline{s}+\ldots+c_{m+1}u_{m+1}\underline{s} = \underline{0}\ . \tag{4.3.10}$$

---

[†] Heft [24].

This implies that the (N+1)-vectors corresponding to $u_1\underline{s}, u_2\underline{s}, \ldots, u_{m+1}\underline{s}$ are dependent. (The zero vector of $V_\theta$ corresponds to the zero vector of $V_{N+1}$.) Hence there are at most (m+1) independent (N+1)-vectors corresponding to $W_\theta(\underline{s})$.

Consider the set of (N+1)-vectors corresponding to the θ-vectors $\underline{s}, \eta\,\underline{s}, \ldots, \eta^m\underline{s}$ of $W_\theta(\underline{s})$, where $\eta$ is a primitive element of $GF(q^{m+1})$. If these (N+1)-vectors are dependent, then there exist $c_0, c_1, \ldots, c_m$ in GF(q), not all zero, such that

$$c_0\underline{s}+c_1\eta\,\underline{s}+\ldots+c_m\eta^m\underline{s} = \underline{0}\ . \tag{4.3.11}$$

This implies that $\eta$ satisfies a polynomial of degree less than (m+1). But this is impossible.

Hence the $q^{m+1}$ (N+1)-vectors of $V_{N+1}$ corresponding to $q^{m+1}$ vectors of $W_\theta(\underline{s})$, all belong to the same (m+1)-dimensional subspace of $V_{N+1}$.

Lemma (4.3.3)[†]. The (N+1)-vectors of $V_{N+1}$ over GF(q) corresponding to all the θ-vectors of $V_\theta$ over $GF(q^{m+1})$ belonging to the k-dimensional subspace of $V_\theta$ are the elements of a k(m+1)-dimensional subspace of $V_{N+1}$, $1 \le k \le \theta$. Also if $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_k$ are k independent θ-vectors of $V_\theta$ then the (N+1)-vectors corresponding to

$$\underline{s}_1,\ \eta\,\underline{s}_1, \ldots,\ \eta^m\underline{s}_1, \underline{s}_2, \eta\,\underline{s}_2, \ldots, \eta^m\underline{s}_k$$

are k(m+1) independent (N+1)-vectors of $V_{N+1}$, where $\eta$ is a primitive element of $GF(q^{m+1})$.

Proof. Suppose the (N+1)-vectors of $V_{N+1}$ corresponding to $\underline{s}_1$, $\eta\,\underline{s}_1$, $\eta^2\underline{s}_1, \ldots, \eta^m\underline{s}_k$ are dependent. Then, without loss of generality, suppose the (N+1)-vector corresponding to $\eta^e\underline{s}_h$ is dependent on the

---

[†] Heft [24]

remaining $\{k(m+1)-1\}$ $(N+1)$-vectors corresponding to $\eta^i\underline{s}_j$, $i \neq e$ and $j \neq h$. Hence there exist $a_{ij}$ in $GF(q)$, not all zero, such that

$$\eta^e\underline{s}_h = \sum_{\substack{0\\ i\neq e}}^{m} \; \sum_{\substack{1\\ j\neq h}}^{k} a_{ij}\,\eta^i\underline{s}_j \tag{4.3.12}$$

or,

$$(\eta^e - \sum_{i\neq e} a_{ih}\eta^i)\underline{s}_h = \sum_{j\neq h} (\sum_{i=0}^{m} a_{ij}\eta^i)\underline{s}_j \, . \tag{4.3.13}$$

But $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_k$ are independent. So we have

$$(\eta^e - \sum_{i\neq e} a_{ih}\eta^i)\underline{s}_h = \underline{0} \, . \tag{4.3.14}$$

That is, $\eta$ satisfies a non-zero polynomial of degree less than $(m+1)$, which is impossible. Hence the $k(m+1)$ $(N+1)$-vectors corresponding to $\eta^i\underline{s}_j$, $i = 0,1,2,\ldots,m$; $j = 1,2,\ldots,k$ are independent.

Let $W_k$ be a k-dimensional subspace in $V_\theta$ over $GF(q^{m+1})$. Let $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_k$ be $k$ independent $\theta$-vectors in $W_k$. Then any $\theta$-vector $\underline{s}$ of $W_k$ can be expressed as

$$\underline{s} = \sum_{i=1}^{k} u_i\underline{s}_i \, , \tag{4.3.15}$$

where $u_1, u_2, \ldots, u_k$ are elements of $GF(q^{m+1})$. There exist , for any $u_i$, elements $a_{i0}, a_{i1}, \ldots, a_{im}$ in $GF(q)$ such that

$$u_i = \sum_{j=0}^{m} a_{ij}\eta^j, \qquad i = 1,2,\ldots,k. \tag{4.3.16}$$

$\eta$ is a primitive element of $GF(q^{m+1})$. So, from (4.3.15), we have

$$\underline{s} = \sum_{i=1}^{k} \sum_{j=0}^{m} a_{ij}\eta^j \, \underline{s}_i \, . \tag{4.3.17}$$

That is, every $(N+1)$-vector in $V_{N+1}$ corresponding to a $\theta$-vector in $W_k$ is dependent on the $(N+1)$-vectors in $V_{N+1}$ corresponding to $\eta^i\underline{s}_j$,

$i = 0,1,2,\ldots,m$; $j = 1,2,\ldots,k$. The number of such (N+1)-vectors is equal to the number of $\theta$-vectors in $W_k$; i.e. $q^{k(m+1)}$. But this is the number of (N+1)-vectors in a $k(m+1)$-dimensional subspace. Hence the result.

## 4.4 Spreads.

A *spread* $S_m$ of m-flat spaces in PG(N,q) is a set of m-flat spaces such that every point in PG(N,q) is contained in exactly one element of $S_m$.

Spreads have been studied by various authors - Bose and Barlotti [8], Bruck [13], Bruck and Bose [14] and [15], Rao [29], Dembowski [20], and Segrê [32].

One of the fundamental results for the existence of spreads is

Theorem (4.4.1). A necessary and sufficient condition for the existence of a spread of m-flat spaces in PG(N,q) is that (m+1) divides (N+1).

Proof. Suppose there exists a spread $S_m$ of m-flat spaces. Then $(q^{m+1}-1)/(q-1)$ divides $(q^{N+1}-1)/(q-1)$. But the necessary and sufficient condition for $(q^{m+1}-1)/(q-1)$ to divide $(q^{N+1}-1)/(q-1)$ is that (m+1) divides (N+1).

We shall prove the sufficient condition by constructing a spread $S_m$. Let $(N+1) = \theta(m+1)$, where N, m, $\theta(>1)$ are positive integers. Consider any point P in $PG(\theta-1, q^{m+1})$. Then P corresponds to a 1-dimensional subspace $W_1$, of a $\theta$-dimensional vector space $V_\theta$ over $GF(q^{m+1})$. Let $\underline{s}$ be a basis of $W_1$. Then

$$W_1 = \{c\ \underline{s}:\ c \in GF(q^{m+1})\}.$$

By lemma (4.3.2), the (N+1)-vectors over GF(q) corresponding to $W_1$ form a (m+1)-dimensional subspace of $V_{N+1}$. But, by lemma (4.3.1), there is a one-one correspondence between m-flat spaces of PG(N,q) and the (m+1)-dimensional subspaces of $V_{N+1}$. So P is represented by an m-flat space in PG(N,q). But P is an arbitrary point. Hence every point of $PG(\theta-1,q^{m+1})$ is represented by an m-flat space of PG(N,q).

Let $\underline{s}_1$ and $\underline{s}_2$ be any two points in $PG(\theta-1,q^{m+1})$. Then let

$$W_\theta(\underline{s}_1) = \{c\ \underline{s}_1: \quad c \in GF(q^{m+1})\}$$

and

$$W_\theta(\underline{s}_2) = \{c\ \underline{s}_2: \quad c \in GF(q^{m+1})\}$$

be the corresponding 1-dimensional subspaces in $V_\theta$. Let $S_i$ (i = 1,2) be the corresponding (m+1)-dimensional subspace of $W_\theta(\underline{s}_i)$ (i = 1,2) in $V_{N+1}$. Let $\sum_m^1$ and $\sum_m^2$ be the m-flat PG(N,q). That is,

$$\underline{s}_1 \longleftrightarrow W_\theta(\underline{s}_1) \longrightarrow S_1 \longleftrightarrow \textstyle\sum_m^1 ,$$

and

$$\underline{s}_2 \longleftrightarrow W_\theta(\underline{s}_2) \longrightarrow S_2 \longleftrightarrow \textstyle\sum_m^2$$

Suppose $\sum_m^1$ and $\sum_m^2$ have a common point, say x. That is

$$\textstyle\sum_m^1 \cap \sum_m^2 = \{x\} .$$

Then the 1-dimensional vector space corresponding to x, say D(x), belongs to both (m+1)-dimensional subspaces $S_1$ and $S_2$ of $V_{N+1}$. That is,

$$S_i \supset D(x), \qquad i = 1,2.$$

But there is one-one correspondence between the points of $V_\theta$ over $GF(q^{m+1})$ and the points of $V_{N+1}$ over GF(q). So the q θ-vectors corresponding to the q (N+1)-vectors of D(x) can be written as either

$$\underline{0},\ c_1u_1\underline{s}_1,\ c_2u_1\underline{s}_1,\ \ldots,\ c_{q-1}u_1\underline{s}_1,$$

or

$$\underline{0},\ c_1u_2\underline{s}_2,\ c_2u_2\underline{s}_2,\ \ldots,\ c_{q-1}u_2\underline{s}_2,$$

where $u_1$ and $u_2$ are non-zero elements of $GF(q^{m+1})$; and $c_i$ $(i = 1,2,\ldots,q-1)$ are non-zero elements of $GF(q)$. Then

$$u_1\underline{s}_1 = u_2\underline{s}_2 .$$

or

$$\underline{s}_1 = u_1^{-1}\ u_2\ \underline{s}_2$$

So, $\underline{s}_1$ and $\underline{s}_2$ represent the same point in $PG(\theta-1,\ q^{m+1})$. Hence the m-flats corresponding to these points are equal, $\sum_m^1 = \sum_m^2$ .

Therefore the m-flat spaces of $PG(N,q)$ representing the points of $PG(\theta-1,\ q^{m+1})$ are disjoint. The number of points in $PG(\theta-1,\ q^{m+1})$ is

$$\frac{q^{\theta(m+1)}-1}{q^{m+1}-1} = \frac{(q^{N+1}-1)}{(q^{m+1}-1)},$$

and the number of points in an m-flat space is

$$(q^{m+1}-1)/(q-1).$$

So all the $(q^{N+1}-1)/(q-1)$ points of $PG(N,q)$ are accounted for. That is, $S_m$ is the set of m-flat spaces corresponding to the points of $PG(\theta-1,\ q^{m+1})$.

Example (4.4.1). Let $q = 2$, $N = 3$ and $m = 1$.

Then $\theta = 2$.

The elements of $GF(2^2)$ are

$$0,\ x,\ x^2,\ 1$$

where

$$x^2 + x + 1 = 0.$$

The points of $PG(1,2^2)$ are

$$\begin{array}{llll} P_1: & \underline{s}_1 = (0,x), & P_4: & \underline{s}_4 = (1,\ 1), \\ P_2: & \underline{s}_2 = (1,x^2), & P_5: & \underline{s}_5 = (x,0)\ , \\ P_3: & \underline{s}_3 = (1,x). & & \end{array}$$

$$\begin{array}{llllllll} W(\underline{s}_1): & 00 & 0x^2 & 01 & 0x & \longleftrightarrow & P_1 \\ W(\underline{s}_2): & 00 & x1 & x^2x & 1x^2 & \longleftrightarrow & P_2 \\ W(\underline{s}_3): & 00 & xx^2 & x^21 & 1x & \longleftrightarrow & P_3 \\ W(\underline{s}_4): & 00 & xx & x^2x^2 & 11 & \longleftrightarrow & P_4 \\ W(\underline{s}_5): & 00 & x^20 & 10 & x0 & \longleftrightarrow & P_5 \end{array}$$

The 2-dimensional subspaces corresponding to $W(\underline{s}_i)$ are:

| | | | | |
|---|---|---|---|---|
| $S_1$: | 0000 | 0011 | 0001 | 0010 |
| $S_2$: | 0000 | 1001 | 1110 | 0111 |
| $S_3$: | 0000 | 1011 | 1101 | 0110 |
| $S_4$: | 0000 | 1010 | 1111 | 0101 |
| $S_5$: | 0000 | 1100 | 0100 | 1000 |

Finally, the 1-flat spaces in $PG(3,2)$ are

| 1-flat spaces | points | | |
|---|---|---|---|
| $L_1$: | 0011 | 0001 | 0010 |
| $L_2$: | 1001 | 1110 | 0111 |
| $L_3$: | 1011 | 1101 | 0110 |
| $L_4$: | 1010 | 1111 | 0101 |
| $L_5$: | 1100 | 0100 | 1000 . |

A spread $S_1$, of 1-flat spaces in $PG(3,2)$ is

(i) $x_0 = 0,\ x_1 = 0;$

(ii) $x_1+x_2 = 0,\ \ x_0+x_2+x_3 = 0;$

(iii) $x_0+x_3 = 0,\ \ x_1+x_2+x_3 = 0;$

(iv) $x_0+x_2 = 0,\ \ x_1+x_3 = 0;$

(v) $x_2 = 0,\ \ x_3 = 0.$

## 4.5 A method for obtaining (N×m) matrices.

Let

$$H = ((h_{ij}))\ ,\quad i = 1,2,\ldots,N;\ \ j = 1,2,\ldots,m,$$

where $h_{ij} \in GF(q)$, be an $(N\times m)$ matrix with rank $m$ and property $R_t(r_1,r_2,\ldots,r_\ell)$. Let $H_1,H_2,\ldots,H_\ell$ be disjoint submatrices (that is no rows in common) of $H$ such that

$$H = \begin{bmatrix} H_1 \\ \cdots\cdots \\ H_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_\ell \end{bmatrix} \tag{4.5.1}$$

$H_i$ is a $(r_i\times m)$ matrix, $i = 1,2,\ldots,\ell$. Then

$$N = \sum_{i=1}^{\ell} r_i \tag{4.5.2}$$

Also

$$\text{rank } H_i = r_i\ ,\quad i = 1,2,\ldots,\ell, \tag{4.5.3}$$

as $H$ has property $R_t(r_1,r_2,\ldots,r_\ell)$. Further, if $(h_{i1},h_{i2},\ldots,h_{im})$ is any row in $H$, then the m-vectors

$$(c\ h_{i1},\ c\ h_{i2},\ldots,c\ h_{im}),\quad c \in GF(q),$$

do not occur in the matrix $H$. That is, the row vectors of $H$ are distinct points of PG(m-1),q). So, for any i, the $r_i$ points of PG(m-1,q) corresponding to the $r_i$ row vectors of $H_i$, are independent. So they generate a $(r_i-1)$-flat space in PG(m-1,q). That is, the $r_i$ row vectors of $H_i$ correspond to a basis of a $(r_i-1)$-flat space in PG(m-1,q). Let $\sum_{r_1-1}, \sum_{r_2-1}, \ldots, \sum_{r_\ell-1}$ be the $\ell$ flat spaces corresponding to the submatrices $H_1,H_2,\ldots,H_\ell$. Let $\sum_{r_{i_1}-1},\sum_{r_{i_2}-1},\ldots,\sum_{r_{i_t}-1}$ be any t of these flat spaces. Then their join is a $\{(\sum_{j=1}^{t} r_{i_j})-1\}$-flat space in PG(m-1,q), since the rank of the submatrix

$$G = \begin{bmatrix} H_{i_1} \\ \cdots\cdots \\ H_{i_2} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_{i_t} \end{bmatrix} \qquad (4.5.4)$$

is $(\sum_{j=1}^{t} r_{i_j})$.

So to construct an (N×m) matrix H with property $R_t(r_1,r_2,\ldots,r_\ell)$, it is sufficient to obtain disjoint flat spaces $\sum_{r_1-1},\sum_{r_2-1},\ldots,\sum_{r_\ell-1}$ of PG(m-1,q) such that the join of any t of these flat spaces, $\sum_{r_{i_1}-1}, \sum_{r_{i_2}-1}, \ldots, \sum_{r_{i_t}-1}$ is a $\{(\sum_{j=1}^{t} r_{i_j})-1\}$-flat space.

**Theorem (4.5.1).** There exists an (N×m) matrix H over GF(q) with property $R_2(\ell\times n)$ if n divides m, where $N = \ell\times n$, $\ell = (q^m-1) \div (q^n-1)$.

Proof. Suppose n divides m. Then $\ell = (q^m-1)/(q^n-1)$ is an integer. Let $\pi_1, \pi_2, \ldots, \pi_\ell$ be a spread $S_{n-1}$ of (n-1)-flat spaces in PG(m-1,q). Then by theorem (4.3.1),

$$|S_{n-1}| = \ell = (q^m-1)/(q^n-1). \tag{4.5.5}$$

Without loss of generality, we can take for any i,

$$\begin{Bmatrix} h_{(i-1)n+1,1}, & h_{(i-1)n+1,2}, & \ldots, & h_{(i-1)n+1,m} \\ h_{(i-1)n+2,1}, & h_{(i-1)n+2,2}, & \ldots, & h_{(i-1)n+2,m} \\ \vdots & \vdots & & \vdots \\ h_{in,1}, & h_{in,2}, & & h_{in,m} \end{Bmatrix} \tag{4.5.6}$$

as a set of n independent points in the flat space $\pi_i$. Then the (n×m) matrix $H_i$ formed by taking the m-vectors of (4.5.6) as row vectors, has rank n. Further, for any two (n×m) matrices $H_i$ and $H_j$, the matrix

$$G = \begin{bmatrix} H_i \\ \cdots\cdots \\ H_j \end{bmatrix}$$

has rank 2n, since the join of $\pi_i$ and $\pi_j$ is a (2n-1)-flat space. Hence the matrix

$$\underset{\ell n \times m}{H} = \begin{bmatrix} H_1 \\ \cdots\cdots \\ H_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_\ell \end{bmatrix}$$

has property $R_2(\ell \times n)$.

Definition (4.5.1). An $(N \times m)$ matrix $H$ of rank $m$ with property $R_t(r_1, r_2, \ldots, r_\ell)$ is said to be complete if it is not possible to adjoin an $(r_{\ell+1} \times m)$ submatrix $H_{\ell+1}$ to $H$ such that

$$\underset{\left(\sum_{i=1}^{\ell+1} r_i\right) \times m}{H^*} = \begin{bmatrix} H \\ \cdots\cdots \\ H_{\ell+1} \end{bmatrix} \quad , \; r_{\ell+1} \geq 1$$

has property $R_t(r_1, r_2, \ldots, r_\ell, r_{\ell+1})$.

Lemma (4.5.2). The $(\ell n \times m)$ matrix $H$ obtained in theorem (4.5.1) is complete.

Proof. The $\ell$ $(n-1)$-flat spaces, $\pi_1, \pi_2, \ldots, \pi_\ell$ corresponding to $H_1, H_2, \ldots, H_\ell$ of $H$ form a spread of $PG(m-1,q)$. Hence any $m$-vector in $PG(m-1,q)$ is covered by one of $(n-1)$-flat spaces, $\pi_1, \pi_2, \ldots, \pi_\ell$. That is any $m$-vector not in $H$ will be dependent on the rows of $H_1, H_2, \ldots, H_\ell$. Hence the result.

Example (4.5.1). $q = 2$, $n = 2$, $m = 4$.

So

$$\ell = \frac{2^4 - 1}{2^2 - 1} = 5 \; .$$

From example (4.4.1), the spread $S_1$ of 1-flat spaces in $PG(3,2)$ is

| | |
|---|---|
| $x_0 = 0, \; x_1 = 0$ | {0011, 0001, 0010} |
| $x_2 = 0, \; x_3 = 0$ | {1100, 0100, 1000} |
| $x_0 + x_2 = 0, \; x_1 + x_3 = 0$ | {1010, 0101, 1111} |
| $x_0 + x_3 = 0, \; x_1 + x_2 + x_3 = 0$ | {1011, 1101, 0110} |
| $x_1 + x_2 = 0, \; x_0 + x_2 + x_3 = 0$ | {1001, 1110, 0111} |

Hence the matrix H with property $R_2(5\times 2)$ is

$$\underset{10\times 4}{H} = \begin{bmatrix} 0001 \\ 0010 \\ 0100 \\ 1000 \\ 1010 \\ 0101 \\ 1011 \\ 1101 \\ 1110 \\ 0111 \end{bmatrix}$$

Notice that it has rank 4. But if we take any 4 rows of H then they may not be independent. For example,

$$\begin{matrix} 0001 \\ 0010 \\ 0100 \\ 0111 \end{matrix}$$

are dependent. Also any three rows may not be independent. For example,

$$\begin{matrix} 0001 \\ 0100 \\ 0101. \end{matrix}$$

Suppose H is a matrix with property $R_2(s_1r_1, s_2r_2, \ldots, s_kr_k)$ and rank m. That is there are $s_i$ $(r_i\times m)$ submatrices $(i = 1,2,\ldots,k)$. Also

$$\ell = s_1+s_2+\ldots+s_k$$

and

$$N = \sum_{i=1}^{k} s_ir_i \quad .$$

Theorem (4.5.3). There exists an $(N\times m)$ matrix H with property $R_2(s_1r_1, s_2r_2, \ldots, s_kr_k)$ if

i) $r_1$ divides m ,

ii) $r_i$ divides $r_{i-1}$, $i = 2,3,\ldots,k$,

where

$$r_1 > r_2 > \dots > r_k$$

and

$$0 \le s_i \le \frac{q^m-1}{q^{r_i}-1} - \sum_{j=1}^{i-1} s_j \frac{q^{r_j}-1}{q^{r_i}-1} . \tag{4.5.7}$$

and

$$s_k = \frac{q^m-1}{q^{r_k}-1} - \sum_{j=1}^{k-1} s_j \frac{q^{r_j}-1}{q^{r_k}-1} . \tag{4.5.8}$$

Proof. Let $S^1_{r_1-1} = \{ \Sigma^1_{r_1-1}, \Sigma^2_{r_1-1}, \dots, \Sigma^{h_1}_{r_1-1}\}$ be a spread of $(r_1-1)$-flat spaces of PG(m-1,q), where $h_1 = (q^m-1)/(q^{r_1}-1)$. The spread $S^1_{r_1-1}$ exists as $r_1$ divides m. Now take $H_1, H_2, \dots, H_{s_1}$ as the $(r_1 \times m)$ matrices of rank $r_1$ corresponding to $\Sigma^1_{r_1-1}, \Sigma^2_{r_1-1}, \dots, \Sigma^{s_1}_{r_1-1}$. Delete these $(r_1-1)$-flat spaces from $S^1_{r_1-1}$. Then each of the $(r_1-1)$-flat spaces in the deleted set has a spread of $(r_2-1)$-flat spaces as $r_2$ divides $r_1$. Let $S^2_{r_2-1}$ denote the collection of $(r_2-1)$-flat spaces of the spreads corresponding to the $(r_1-1)$-flat spaces, $\Sigma^{s_1+1}_{r_1-1}, \Sigma^{s_1+2}_{r_1-1}, \dots, \Sigma^{h_1}_{r_1-1}$. Without loss of generality, we can take

$$S^2_{r_2-1} = \{ \Sigma^1_{r_2-1}, \Sigma^2_{r_2-1}, \dots, \Sigma^{h_2}_{r_2-1}\} ,$$

where

$$h_2 = \left(\frac{q^m-1}{q^{r_1}-1} - s_1\right) \times \frac{q^{r_1}-1}{q^{r_2}-1} . \tag{4.5.9}$$

Now take $H_{s_1+1}, H_{s_1+2}, \dots, H_{s_1+s_2}$ as the $(r_2 \times m)$ matrices of rank $r_2$ corresponding to the first $s_2$ $(r_2-1)$-flat spaces in $S^2_{r_2-1}$. Similarly, the rest of the submatrices of rank $r_3, r_4, \dots, r_k$, can be obtained.

Clearly the matrix H so obtained has the property $R_2(s_1r_1, s_2r_2, \ldots, s_kr_k)$ as we are dealing with submatrices corresponding to disjoint flat spaces.

The matrix so obtained is complete as each point of the geometry PG(m,q) is covered exactly once.

<u>Example (4.5.2)</u>. $q = 2,\ m = 4,\ r_1 = 2,\ r_2 = 1,\ s_1 = 3,\ s_2 = 6$.

So $\ell = 9$ and $N = 12$.

$$h_1 = \frac{2^4-1}{2^2-1} = 5$$

So, $0 \le s_1 \le 5$

and

$$h_2 = (5-s_1)\times 3 = 15-3s_1\ .$$

We shall construct a matrix H with property $R_2(3\times 2,\ 6\times 1)$.

$$S_1^1 = \begin{cases} x_0 = 0, & x_1 = 0 \\ x_2 = 0, & x_3 = 0 \\ x_0+x_2 = 0, & x_1+x_3 = 0 \\ x_0+x_3 = 0, & x_1+x_2+x_3 = 0 \\ x_1+x_2 = 0, & x_0+x_2+x_3 = 0 \end{cases}$$

So

$$H_1 = \begin{bmatrix} 0001 \\ 0010 \end{bmatrix}$$

$$H_2 = \begin{bmatrix} 0100 \\ 1000 \end{bmatrix}$$

$$H_3 = \begin{bmatrix} 0101 \\ 1010 \end{bmatrix}$$

From the spread of $\{x_0+x_3=0,\ x_1+x_2+x_3=0\}$, we have

$H_4 = [1011]$,

$H_5 = [1101]$,

$H_6 = [0110]$.

Also from the spread of the line, $\{x_1+x_2=0,\ x_0+x_2+x_3=0\}$ we have

$H_7 = [1001]$,

$H_8 = [1110]$,

$H_9 = [0111]$.

so, the desired matrix H with rank 4 is

$$\underset{12\times 4}{H} = \begin{bmatrix} 0001 \\ 0010 \\ 0100 \\ 1000 \\ 0101 \\ 1010 \\ 1011 \\ 1101 \\ 0110 \\ 1001 \\ 1110 \\ 0111 \end{bmatrix} \qquad (4.5.10)$$

We can also obtain matrices H when n does not divide m, with property $R_2(\ell \times n)$. But the matrices so obtained may not be complete.

Example (4.5.3). $q = 2$, $t = 2$, $n = 2$ and $m = 5$.

Consider the geometry PG(4,2). It has 31 points. There does not exist a spread of lines (1-flat spaces) in the geometry. But consider the following 1-flat spaces and their points

$$\begin{array}{llll} L_1: & 01001 & 10010 & 11011 \\ L_2: & 10001 & 11010 & 01011 \\ L_3: & 11001 & 01010 & 10011 \\ L_4: & 00001 & 01101 & 01100 \\ L_5: & 00010 & 10101 & 10111 \\ L_6: & 00011 & 11101 & 11110 \\ L_7: & 10000 & 00100 & 10100 \\ L_8: & 01000 & 00110 & 01110 \\ L_9: & 11000 & 00111 & 11111 \end{array} \tag{4.5.11}$$

These nine lines are disjoint and cover 27 points of the geometry. The remaining four points are

$$P_{28}:\ 00101,\ P_{29}:\ 01111;\ P_{30}:\ 10110;\ P_{31}:\ 11100 \tag{4.5.12}$$

Any line passing through any one of these four points intersects at least one of the above (4.5.11) nine lines. Now consider the matrix $H$ obtained from the above nine 1-flats,

$$H = \begin{bmatrix} 01001 \\ 10010 \\ 10001 \\ 11010 \\ 11001 \\ 01010 \\ 00001 \\ 01101 \\ 00010 \\ 10101 \\ 00011 \\ 11101 \\ 10000 \\ 00100 \\ 01000 \\ 00110 \\ 11000 \\ 00111 \end{bmatrix} . \tag{4.5.13}$$

The matrix $H$ clearly has property $R_2(9\times2)$. But it is not complete as we can adjoin any one of four points (4.5.12) to have property $R_2(9\times2,1)$.

This result can be extended for higher values of m, when n does not divide m. We shall illustrate this method by an example.

Example (4.5.4). $m = 7$, $n = 2$, $q = 2$, $t = 2$.

Adjoin "01" to all the odd rows of the matrix in (4.5.13) and "10" to all the even rows. For example,

$$\begin{matrix} 0101001 \\ 1010010 \\ 0110001 \\ 1011010 \end{matrix}$$

and so on. Let $G_1$ be the resulting (18×7) matrix. Clearly $G_1$ has property $R_2(9\times 2)$. We can also obtain (18×7) matrices from the matrix (4.5.13) by adjoining (10,11); (11,01) and (00,00). Let $G_2$, $G_3$ and $G_4$ be the resulting matrices. Each of these matrices $G_2$, $G_3$ and $G_4$ have property $R_2(9\times 2)$.

Consider the first two rows of $G_1$ and the first two rows of $G_2$, i.e.

$$\begin{matrix} 0101001 & 1001001 \\ 1010010 & 1110010 \end{matrix}$$

These four points are independent. That is the 1-flat passing through the points 0101001 and 1010010 does not intersect the 1-flat passing through 1001001 and 1110010. Similarly we can show that the remaining pairs are independent. Hence the matrix,

$$\begin{bmatrix} G_1 \\ \cdots\cdots \\ G_2 \\ \cdots\cdots \\ G_3 \\ \cdots\cdots \\ G_4 \end{bmatrix} \qquad (4.5.14)$$

obtained by adjoining $G_1, G_2, G_3$ and $G_4$, has property $R_2(36\times 2)$. Consider the lines

$$\{1000000,\ 0000101,\ 1000101\}\ ,$$
$$\{0100000,\ 0001111,\ 0101111\}\ ,$$
$$\{1100000,\ 0010110,\ 1110110\}.$$

These lines do not intersect any one of the lines corresponding to the matrix in (4.5.14), as the points 00101, 01111, and 10110 do not lie on the 1-flat spaces, given in (4.5.11). Let

$$G_5 = \begin{bmatrix} 1000000 \\ 0000101 \\ 0100000 \\ 0001111 \\ 1100000 \\ 0010110 \end{bmatrix} \qquad (4.5.15)$$

Then the matrix

$$\underset{39\times 7}{H} = \begin{bmatrix} G_1 \\ \cdots\cdots \\ G_2 \\ \cdots\cdots \\ G_3 \\ \cdots\cdots \\ G_4 \\ \cdots\cdots \\ G_5 \end{bmatrix} \qquad (4.5.16)$$

has property $R_2(39\times 2)$. But it is not complete, as the corresponding lines do not form a spread of 1-flat spaces in PG(6,2). Also we cannot adjoin more than one point to $H$ in (4.5.16).

## 4.6 General methods

Consider the matrix,

$$\underset{(12\times 6)}{H} = \begin{bmatrix} 000001 \\ 000010 \\ 000100 \\ 001000 \\ 010000 \\ 100000 \\ 010101 \\ 101010 \\ 100111 \\ 111001 \\ 011011 \\ 110110 \end{bmatrix} \tag{4.6.1}$$

It has property $R_3(6\times 2)$ and is complete. But the construction of $(N\times m)$ matrices with property $R_t(r_1,r_2,\ldots,r_\ell)$, $t \geq 3$, which are complete, is a very difficult problem. We shall give a simple procedure for obtaining $(N\times m)$ matrices with the desired property. But these matrices may not be complete.

Let

$$r = \max(r_1,r_2,\ldots,r_\ell).$$

An $(N\times m)$ matrix with rank $m$, is said to have the property $P_t$ if no $t$ rows of the matrix are dependent.

Lemma (4.6.1). If $H$ is an $(N\times m)$ matrix with property $P_{tr}$ then $H$ has property $R_t(r_1,r_2,\ldots,r_\ell)$.

Proof. $H$ is a matrix with property $P_{tr}$. We can express $H$ as

$$H = \begin{bmatrix} H_1 \\ \cdots\cdots \\ H_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_\ell \end{bmatrix} \tag{4.6.2}$$

where $H_i$ is a $(r_i \times m)$ submatrix and $\sum_{i=1}^{\ell} r_i = N$. Consider any $t$ submatrices - say $H_{i_1}, H_{i_2}, \ldots, H_{i_t}$. Let

$$G = \begin{bmatrix} H_{i_1} \\ \cdots\cdots \\ H_{i_2} \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ H_{i_t} \end{bmatrix} \tag{4.6.3}$$

Then the number of rows in $G$ is equal to $\sum_{j=1}^{\ell} r_{i_j}$. But

$$\sum_{j=1}^{\ell} r_{i_j} \leq t \times r .$$

Hence

$$\text{rank } G = \sum_{j=1}^{\ell} r_{i_j} .$$

That is, H has property $R_t(r_1, r_2, \ldots, r_\ell)$. But the converse of this result is not necessarily true.

Example (4.6.1). Consider the (10×4) matrix in the example (4.5.1). It has property $R_2(5\times2)$ but any four rows are not independent.

Now we shall describe, in brief, a method for obtaining $(N\times m)$ matrices with elements from $GF(q)$, which are of rank m and have property $P_t$. These matrices were first obtained by Bose [5] in connection with the problem of confounding symmetrical factorial designs.

Let $t < N \leq q^\theta - 1$, where $t$, $N$ and $\theta$ are integers. Consider the extension field $GF(q^\theta)$ of $GF(q)$. We can obtain a one-one correspondence between the elements of $GF(q^\theta)$ and $\theta$-vectors of the vector space $V_\theta$, as in the preceeding section (4.3),

$$(a_0, a_1, \ldots, a_{\theta-1}) \longleftrightarrow a_0 + a_1\delta + \ldots + a_{\theta-1}\delta^{\theta-1}$$

where $\delta$ is a primitive element of $GF(q^\theta)$. Let $\delta_1, \delta_2, \ldots, \delta_N$ be the distinct non-zero elements of $GF(q^\theta)$. In particular we can choose

$$\delta_i = \delta^{i-1}, \qquad i = 1,2,\ldots,N .$$

Let

$$H_0 = \begin{bmatrix} \delta_1 & \delta_1^2 & \delta_1^3 & \ldots & \delta_1^t \\ \delta_2 & \delta_2^2 & \delta_2^3 & \ldots & \delta_2^t \\ \vdots & \vdots & \vdots & & \vdots \\ \delta_N & \delta_N^2 & \delta_N^3 & \ldots & \delta_N^t \end{bmatrix} \qquad (4.6.4)$$

The determinant of any submatrix of $H_0$, formed by $i_1$-th, $i_2$-th, ..., $i_t$-th rows is

$$\prod (\delta_{i_u} - \delta_{i_v}) , \qquad u, v = 1,2,\ldots,t;\ u \neq v$$

and is therefore non-zero. Hence no $t$ rows of $H_0$ are dependent.

Since $x \to x^q$, is an automorphism of $GF(q^\theta)$, and $\eta^q = \eta$, if $\eta$ is an element of $GF(q)$, it follows that if we delete from $H_0$ the columns headed by $\delta_1^q$, $\delta_1^{2q}$, ..., $\delta_1^{sq}$ where $s = [\frac{t}{q}]^\dagger$ then any $t$ rows of the resulting matrix $H_1$ will still be independent over $GF(q)$; i.e. no linear combination of $t$ rows of $H_1$, with coefficients from $GF(q)$ will vanish. If we now regard the elements of $H_1$ as row vectors of $V_\theta$ we have a $N \times \theta(t - [\frac{t}{q}])$ matrix with elements from $GF(q)$, which has property $P_t$. If there are any columns in $H_1$ that are dependent on others we can delete them and still obtain a matrix $H$ with property

---

$^\dagger$[x] indicates the largest integer not exceeding x.

$P_t$. Let the rank of the matrix H be m. Hence

Theorem (4.6.2). If $t < N \leq q^{\theta}-1$, then we can find an $(N \times m)$ matrix H of rank m, with elements from GF(q), and having property $P_t$, where $m \leq \theta(t - [\frac{t}{q}])$.

# BIBLIOGRAPHY

[ 1] Abraham, C.T., Ghosh, S.P, and Ray-Chaudhuri, D.K. "File organization schemes based on finite geometries", IBM Research Report RC-1459, Yorktown Heights, New York; IBM Watson Research Center, August 1965.

[ 2] Barlotti, A. "Some topics in finite geometrical structures," Institute of Statistics Mimeo Series No. 439, University of North Carolina, Chapel Hill, August 1965.

[ 3] Bose, R.C. "On the construction of balanced incomplete block designs," Annals of Eugenics, Vol. 9(1939) pp. 353-399.

[ 4] Bose, R.C. "Some new series of balanced incomplete block designs," Bulletin Calcutta Mathematical Society, Vol. 34 (1942) pp. 17-31.

[ 5] Bose, R.C. "Mathematical theory of the symmetrical factorial designs," Sankhya, Vol. 8(1947) pp. 107-166.

[ 6] Bose, R.C. Unpublished notes on Combinatorial Mathematics, University of North Carolina, Chapel Hill.

[ 7] Bose, R.C, Abraham, C.T., and Ghosh, S.P. "File organization of records for multiple-valued attributes for multi-attribute queries," Proceedings of the Symposium on Combinatorial Mathematics, Chapel Hill, North Carolina: University of North Carolina Press, 1967.

[ 8] Bose, R.C. and Barlotti, A. "Linear representation of a class of Projective planes in four dimensional Projective space," Annali Di Mathematica Pura Ed Applicata, Vol.88(1971) pp.9-32.

[ 9] Bose, R.C. and Bush, K.A. "Orthogonal arrays of strength two and three," The Annals of Mathematical Statistics, Vol. 23, No.4 (December 1952), pp. 508-524.

[10] Bose, R.C. and Ray-Chaudhuri, D.K. "On a class of binary error-correcting group codes," Information and Control, Vol. 3 (1960) pp. 68-79.

[11] Bose, R.C. and Ray-Chaudhuri, D.K. "Further results on error-correcting group codes," Information and Control, Vol.3 (1960) pp. 279-298.

[12] Bose, R.C., Shrikhande, S.S. and Bhattacharya, K.N. "On the construction of group divisible incomplete block designs," The Annals of Mathematical Statistics, Vol. 24, No.2 (June 1953) pp. 167-195.

[13] Bruck, R.H. "Construction problem of finite projective planes," Combinatorial Mathematics and its Applications, University of North Carolina Press (1969), pp. 426-514.

[14] Bruck, R.H. and Bose, R.C. "The construction of Translation Planes from Projective Planes," Journal of Algebra, Vol. 1 (1964) pp. 85-102.

[15] Bruck, R.H. and Bose, R.C. "Linear representations of projective planes in projective spaces," Journal of Algebra, Vol. 4 (1966) pp. 117-172.

[16] Buchholz, W. "File organization and addressing," IBM Systems Journal, Vol. 2(1963), pp. 86-111.

[17] Bush, K.A. "Orthogonal arrays of index unity," The Annals of Mathematical Statistics, Vol. 23, No. 4(1952), pp. 426-434.

[18] Chakravarti, I.M. "Fractional replication in asymmetrical factorial designs and partially balanced arrays," Sankhya, Vol. 17(1956), pp. 143-164.

[19] Chakravarti, I.M. "On some methods of construction of partially balanced arrays," The Annals of Mathematical Statistics, Vol. 32, No. 4(1961), pp. 1181-1185.

[20] Dembowski, P. Finite geometries, Springer-Verlag, Berlin, Heidelberg, 1968.

[21] Dowling, T.A. "Two-and three-valued codes for the Gaussian channel," Institute of Statistics Mimeo Series No. 499, University of North Carolina, Chapel Hill, (1966).

[22] Dowling, T.A. "A class of Tri-weight cyclic codes," Institute of Statistics Mimeo Series No. 600.3, University of North Carolina, Chapel Hill, (1969).

[23] Ghosh, S.P. and Abraham, C.T. "Application of finite geometry in file organization for records with multiple valued attributes," IBM Research Report RC-1561, Yorktown Heights, New York, IBM Watson Research Center, March 1966.

[24] Heft, S.M. "Spreads in projective geometry and associated designs," Doctoral dissertation, University of North Carolina, 1971.

[25] Koch, G.G. "The design of combinatorial information retrieval systems for files with multiple-valued attributes," University of North Carolina Mimeo Series No. 552, Chapel Hill, 1967.

[26] Peterson, W.W. Error Correcting Codes, MIT Press, Cambridge, Mass. (1961).

[27] Primrose, E.J.F. "Quadrics in finite geometries," Proceedings Cambridge Philosophical Society, Vol.47(1951), pp. 299-304.

[28] Rao, C.R. "A study of BIB designs with replications 11 to 15," Sankhya, Series A, Vol. 24(1962), pp. 203-207.

[29] Rao, C.R. "Cyclical generation of linear subspaces in finite geometries," Combinatorial Mathematics and its applications, University of North Carolina Press (1969), pp. 515-535.

[30] Ray-Chaudhuri, D.K. "Some results on quadrics in finite projective geometry based on Galois fields," Canadian Journal of Mathematics, Vol. 14(1962), pp. 129-138.

[31] Ray-Chaudhuri, D.K. "Combinatorial information retrieval systems for files," IBM Research Report RC-1554, Yorktown Heights, New York, IBM Watson Research Center, February 1966.

[32] Segre, B. "Teoria di Galois Fibrazioni proiettive e Geometrie Non Desarguesiane," Annali Di Mathematica Pura Ed Applicata, Vol. 64(1964), pp. 1-76.

[33] Sprott, D.A. "A study of BIB designs with replications 16 to 20," Sankhya, Series A, Vol. 24, (1962), pp. 203-207.