GRASSMANNIANS OVER FINITE FIELDS AND MDS CODES

SUDHIR R. GHORPADE AND GILLES LACHAUD

Contents

1. Introduction and Statements of Results

Let n and k be positive integers such that $k \leq n$, and let V be an n–dimensional vector space over $\mathbb{F}_q$, a finite field with q elements. Fix a basis of V so that its elements can be identified with their "co-ordinates". Recall that an (n,k)–linear code over $\mathbb{F}_q$ is simply a k–dimensional subspace of V. Given such a code C, the minimal distance of C is defined by

$$d(C) = \min\{d(x,y) : x, y \in C,\ x \neq y\},$$

where $d(x,y)$ denotes the Hamming distance between x and y, viz., the number of nonzero co-ordinates of $x-y$. The minimal distance always satisfies the "Singleton bound":

$$d(C) \leq n - k + 1.$$

Codes C with $d(C) = n-k+1$ are called *maximal distance separable (MDS)* codes. Since a code with $d(C) = d$ corrects $\left[\dfrac{d-1}{2}\right]$ errors, MDS codes are clearly of considerable interest. Some examples of MDS codes are: binary repetition code with $q = 2$ and $k = 1$, Reed–Soloman codes (which have $n = q-1$), and Goppa codes corresponding to nonsingular projective curves of genus zero having "sufficiently many" $\mathbb{F}_q$–rational points. Some of the earlier results on MDS codes include Singleton's characterization [8] of an (n,k)–MDS code with $k = 2$ as a set of $n-1$ pairwise orthogonal Latin squares, and a result of Goethals [4] showing the existence of MDS codes when $q \geq n > k$. For more on these lines, see, for instance, [2] or [9]. In this note, we consider the more general problem of determining

$$N(n,k;q) \ = \ \text{the number of all } (n,k)\text{–MDS codes over } \mathbb{F}_q.$$

Explicit determination of $N(n,k;q)$ is still an open problem. However, we show that using some elementary techniques in the algebraic geometric study of grassmannians, useful information about $N(n,k;q)$ can be obtained. In greater detail,

---

Date: August 1995.

The first author is partially supported by research grant No. 93-106/RG/MATHS/AS from the Third World Academy of Sciences, Italy.

This is an unofficial electronic version of the paper which has appeared in *Proceedings of the Discussion Meeting on Cryptography and Computation*, (Coordinators: C. E. Veni Madhavan and R. Balasubramanian), Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, 1995.

we show that a lower bound for $N(n,k;q)$ is given by

$$m = \binom{n}{k} q^{k(n-k)} - \left[\binom{n}{k} - 1\right] g_k(n),$$

where $g_k(n)$ is the Gaussian binomial coefficient, given by

$$g_k(n) = \frac{(q^n-1)(q^n-q)\dots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\dots(q^k-q^{k-1})}.$$

We also recover the "obvious" upper bound $M = q^{k(n-k)}$ for $N(n,k;q)$. Further, we show that the lower bound m is asymptotically given by

$$m = q^\delta + \left[1 - \binom{n}{k}\right] \left[q^{\delta-1} + 2q^{\delta-2} + O(q^{\delta-3})\right], \quad \text{where } \delta = k(n-k).$$

This formula may be compared with the following asymptotic result due to G. Lachaud.

$$N(n,k;q) = q^\delta + \left[1 - \binom{n}{k}\right] q^{\delta-1} + O(q^{\delta-2}).$$

Thus it is seen that m is a reasonably good approximation to $N(n,k;q)$ and that the inequality $N(n,k;q) \geq m$ implies the existence of MDS codes for a wide class of parameter values (n,k,q), and, in particular, for all large enough q.

2. Proofs

Let C be an (n,k)–MDS code over $\mathbb{F}_q$ and H be its parity check matrix [i.e., an $(n-k) \times n$ matrix whose rows form a basis of the dual of C in V]. Then it is well-known and easy to see that the minimal distance of C can be characterized as follows.

Lemma 1. *Given any positive integer d, we have $d = d(C)$ if and only if H has d linearly dependent columns and any set of $d-1$ columns of H is linearly independent.* □

Observe that the inequality $d(C) \leq n-k+1$ is an immediate consequence of the above lemma. Now let $r = n-k$ and let $G_{r,n}$ denote the grassmannian consisting of all r–dimensional subspaces of V. It is not difficult to observe (cf. [1] or [5], for example) that

$$|G_{r,n}| = g_r(n) = g_k(n) = \frac{(q^n-1)(q^n-q)\dots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\dots(q^k-q^{k-1})}. \tag{1}$$

It is well-known (see [6] or [7]) that $G_{r,n}$ can be embedded in the projective space $\mathbb{P}^{\binom{n}{k}-1}$ over $\mathbb{F}_q$ via the Plücker embedding. Specifically, the Plücker embedding identifies an r–dimensional subspace W of V with the point $p = (p_\alpha)$ of $\mathbb{P}^{\binom{n}{k}-1}$, indexed by

$$I(r,n) = \{\alpha = (\alpha_1, \dots, \alpha_r) : 1 \leq a_1 < \dots < \alpha_r \leq n\},$$

such that p_α is the $r \times r$ minor of A formed by the columns of A with indices $\alpha_1, \dots, \alpha_r$, where A is an $r \times n$ matrix whose rows give a basis of W. Thus, from the above lemma, it is clear that the set of all (n,k)–MDS codes is in one-to-one correspondence with the set

$$E = \{p = (p_\alpha) \in G_{r,n} : p_\alpha \neq 0, \text{ for all } \alpha \in I(r,n)\}.$$

In order to estimate $|E|$, we shall use a fundamental property of the Plücker coordinates $p = (p_\alpha)$, which is stated below. This essentially appears as Proposition 2 in [7]. It may be noted that although in [7] it is assumed that the ground field is $\mathbb{C}$, the argument therein works in arbitrary characteristic.

Lemma 2. *Fix some $\alpha \in I(r,n)$. Then there is a natural one-to-one correspondence between the points of $\{p \in G_{r,n} : p_\alpha \neq 0\}$ and the points of $k(n-k)$–dimensional affine space (over $\mathbb{F}_q$) of $r \times n$ matrices (t_{ij}) such that the $r \times r$ submatrix $(t_{i\alpha_j})$ is the identity submatrix.*

Corollary. *Given $\alpha \in I(r,n)$, let $E_\alpha = \{p \in G_{r,n} : p_\alpha = 0\}$. Then $|E_\alpha| = g_k(n) - q^{k(n-k)}$.*
Proof. Follows from (1) and Lemma 2. □

Theorem 1. *With m and M as in the Introduction, we have $m \leq N(n,k;q) \leq M$.*
Proof. From Lemma 2, it is clear that $N(n,k;q) = |E| \leq q^{k(n-k)} = M$. Next, using the notation and the conclusion of the above Corollary, we see that

$$\left| \bigcup_{\alpha \in I(r,n)} E_\alpha \right| \leq \sum_{\alpha \in I(r,n)} |E_\alpha| = \binom{n}{k} \left[g_k(n) - q^{k(n-k)} \right].$$

Now since $E = G_{r,n} \setminus \cup_\alpha E_\alpha$, it follows that

$$N(n,k;q) \;=\; |E| \;\geq\; g_k(n) - \binom{n}{k} \left[g_k(n) - q^{k(n-k)} \right] \;=\; m. \qquad \square$$

From classical combinatorics, we recall that $g_k(n)$ is a polynomial in q with positive integral coefficients. In fact,

$$g_k(n) = \sum_{i=0}^{\delta} \gamma_i \, q^i$$

where $\delta = k(n-k)$ and for $0 \leq i \leq \delta$,

$$\begin{aligned} \gamma_i \;&= \text{the number of partitions of } i \text{ into at most } k \text{ parts, each at most } n-k \\ &= \left| \{ (j_1, \ldots, j_s) : j_1 + \cdots + j_s = i, \; s \leq k, \text{ and } n-k \geq j_1 \geq \cdots \geq j_s \geq 1 \} \right|. \end{aligned}$$

Alternately, one can describe these in terms of paths in an $k \times (n-k)$ rectangle. See [1] for details. There is also a topological description (cf. [3, p. 292]), namely, $\gamma_i = \dim H^{2i}(G_{k,n}; \mathbb{C})$. At any rate, we can easily deduce that

$$g_k(n) = q^\delta + q^{\delta-1} + 2q^{\delta-2} + O(q^{\delta-3}).$$

In view of this, Theorem 1 implies the following asymptotic result.

Theorem 2. *For the lower bound m defined in the Introduction, we have*

$$m = q^\delta + \left[1 - \binom{n}{k} \right] \left[q^{\delta-1} + 2q^{\delta-2} + O(q^{\delta-3}) \right], \quad \textit{where } \delta = k(n-k). \qquad \square$$

We conclude with the following

Remark. It is clear that the inequalities we have used in estimating $|E|$ are fairly crude. Yet, the bounds obtained are reasonably good. It seems evident that a deeper analysis of the number of points on the sections of grassmannians by coordinate hyperplanes, should give more precise information about $N(n,k;q)$.

References

[1] G. Andrews, *Theory of Partitions*, Encyclopedia of Mathematics & its Applications, Vol. 2, Addison-Wesley, 1976.

[2] I. F. Blake and R. C. Mullin, *An Introduction to Algebraic and Combinatorial Coding Theory*, Academic Press, 1976.

[3] R. Bott and L. W. Tu, *Differential forms in algebraic topology*, Springer-Verlag, 1980.

[4] J. M. Goethals, *A polynomial approach to linear codes,* Philips Research Reports, **24** (1969), 145–159.

[5] J. Goldman and G.-C. Rota, *The number of subspaces of a vector space*, in: Recent progress in Combinatorics, ed. W. T. Tutte, Academic Press (1969), 75–84.

[6] W. V. D. Hodge and D Pedoe, *Methods of Algebraic Geometry, Vol. II*, Cambridge University Press, 1952.

[7] S. Kleiman and D. Laksov, *Schubert calculus*, American Math. Monthly, **79** (1972), 1061–1082.

[8] R. C. Singleton, *Maximal distance Q-nary codes*, IEEE Trans. Information Theory, **IT-10** (1964), 116–118.

[9] J. H. van Lint, *Introduction to Coding Theory*, Springer-Verlag, 1982.

S.R.G.: Department of Mathematics
Indian Institute of Technology, Bombay
Powai, Mumbai 400076 - INDIA
E-mail address: srg@math.iitb.ac.in

G.L.: Équipe "Arithmétique et Théorie de l'Information"
Institut de Mathématiques de Luminy
Luminy Case 907, 13288 Marseille Cedex 9 - FRANCE
E-mail address: lachaud@iml.univ-mrs.fr