Journal of Combinatorial Theory, Series A **96**, 396–399 (2001)
doi:10.1006/jcta.2001.3206, available online at http://www.idealibrary.com on IDEAL®

NOTE

Linear Codes with Exponentially Many Light Vectors

Alexei Ashikhmin[1] and Alexander Barg[1]

Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974
E-mail: aea@research.bell-labs.com, abarg@research.bell-labs.com

and

Serge Vlăduţ

Institut de Mathématiques de Luminy, UPR 9016 du CNRS, Luminy, Case 907, 13288 Marseille Cedex 9, France
E-mail: vladut@iml.univ-mrs.fr.

Communicated by V. Pless

Received January 3, 2001

G. Kalai and N. Linial (1995, *IEEE Trans. Inform. Theory* **41**, 1467–1472) put forward the following conjecture: Let $\{C_n\}$ be a sequence of binary linear codes of distance d_n and A_{d_n} be the number of vectors of weight d_n in C_n. Then $\log_2 A_{d_n} = o(n)$. We disprove this by constructing a family of linear codes from geometric Goppa codes in which the number of vectors of minimum weight grows exponentially with the length.

1. INTRODUCTION

Let C be a code over $\mathbb{F}_q$ of length n and distance $d = d(C)$. The (Hamming) distance distribution of the code is an $(n+1)$-vector $(A_0 = 1, A_1, ..., A_n)$, where $A_w = A_w(C) := (\sharp C)^{-1} |\{(x, x') \in C^2 : \mathrm{d}(x, x') = w\}|$. Of course $A_w = 0$ if $1 \leqslant w \leqslant d-1$. If C is linear then A_w is the number of vectors of weight w in it.

Let $\{C_{n_i}\}$ be a family of binary linear codes of growing length n_i and let $d_{n_i} = d(C_{n_i})$ (below we omit the subscript i). Kalai and Linial [2] conjectured that for any such family the number A_{d_n} is subexponential in n, i.e., that for any $\alpha > 0$ there is a number $N(\alpha)$ such that for all $n > N(\alpha)$ we

[1] Research supported in part by Binational (USA–Israel) Science Foundation under Grant 1999099.

0097-3165/01 $35.00

have $\log A_{d_n} \leqslant \alpha n$ (if the base of logarithms is missing, it is 2 throughout). They also made a similar conjecture about unrestricted (i.e., not necessarily linear) codes and wrote, "The [asymptotic] distance distribution near the minimum distance remains a great mystery."

While we now know a little more about the distance distribution of codes for larger w [1, 3], this claim is still very much true. The above conjectures, however, are not as will be shown below. Let

$$E_q(\delta) := H(\delta) - \frac{\log q}{\sqrt{q}-1} - \log \frac{q}{q-1},$$

where $H(y) = -y \log y - (1-y) \log(1-y)$. For $q \geqslant 49$ the function $E_q(\delta)$ has two zeros $0 < \delta_1 < \delta_2 < (q-1)/q$ and is positive for $\delta_1 < \delta < \delta_2$.

THEOREM 1. *Let $q = 2^{2s}$, $s = 3, 4, \ldots$ be fixed. Then for any $\delta_1 < \delta < \delta_2$ there exists a sequence of binary linear codes $\{C_n\}$ of length $n = qN$, $N \to \infty$ and distance $d_n = n\delta/2$ such that*

$$\log A_{d_n} \geqslant N E_q(\delta) - o(N). \tag{1}$$

2. PROOF

We will first construct a sequence of q-ary linear (geometric Goppa) codes. Background information on coding theory and geometry of curves can be looked up in [5].

Let X be a (smooth projective absolutely irreducible) curve of genus g over $\mathbb{F}_q$, where $q \geqslant 49$ is an even power of a prime. Let $N = N(X) := \sharp X(\mathbb{F}_q)$ be the number of $\mathbb{F}_q$-rational points of X and suppose that X is such that $N \geqslant g(\sqrt{q}-1)$ (e.g., X is a suitable modular curve). The set of $\mathbb{F}_q$-rational effective divisors of degree $a \geqslant 0$ on X is denoted by $Div_a^+(X)$. Recall that $Div_a^+(X)$ is a finite set. For $D \in Div_a^+(X)$ let $L(D)$ be the corresponding linear system (the linear space of rational functions associated with D). Denote by $\mathscr{C} = \mathscr{C}(D)$ the geometric Goppa code on X defined by the triple $(X, D, X(\mathbb{F}_q))$ in the usual way. $\mathscr{C}$ is a linear code of length N, dimension $\dim(\mathscr{C}) \geqslant a - g + 1$, and distance $d(\mathscr{C}) \geqslant N - a$.

THEOREM 2. *Let $\delta = (N-a)/N$ satisfy the inequality $\delta_1 < \delta < \delta_2$. Then there exists $D \in Div_a^+(X)$ such that the corresponding geometric Goppa code $\mathscr{C} = \mathscr{C}(D)$ has the minimum distance $d = N - a = \delta N$ and for the number A_d of vectors of weight d we have*

$$\log A_d \geqslant N E_q(\delta) - o(N).$$

Proof. The proof follows the ideas of [6]. We set for an integer $r \in [0, a]$

$$C_{a,r} := \left\{ D \in Div_a^+(X) : \sharp \left(\operatorname{Supp} D \bigcap X(\mathbb{F}_q) \right) = r \right\}.$$

We denote by $J_a = J_a(X)$ the set of (linear) classes of degree a divisors. Thus, J_a is the quotient space of $Div_a^+(X)$ under the linear equivalence of divisors. Recall that since $Div_a^+(X)$ is non-empty, J_a is in a bijection with the set $J_X(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points on the Jacobian variety of X.

The following lemma from [4, Lemma A2] (see also [6]) is a key ingredient in the proof.

LEMMA 1.

$$\log_q \sharp J_a = g\left(1 + (\sqrt{q} - 1) \log_q \frac{q}{q-1}\right) - o(g). \tag{2}$$

Further, it is obvious that $\sharp C_{a,a} = \binom{N}{a}$ and so

$$\log \sharp C_{a,a} = NH\left(\frac{a}{N}\right) - o(g). \tag{3}$$

Recall that the fibers of the canonical projection

$$\pi_a \colon Div_a^+(X) \longrightarrow J_a(X)$$

are the projective spaces $\mathbb{P}(D) = \mathbb{P}(L(D))$, which are projectivizations of linear systems $L(D)$. For any $D \in Div_a^+(X)$ the number of words of weight $d = N - a$ in the code $\mathscr{C}(D)$ equals

$$A_d(D) = (q-1)\, \sharp \left(\pi_a^{-1}(\pi_a(D)) \bigcap C_{a,a} \right).$$

Thus we have

$$A_d^* := \max\{A_d(D) : D \in Div_a^+(X)\} \geqslant \frac{\sharp C_{a,a}}{\sharp J_a}.$$

Taking logarithms and using (2), (3) we obtain Theorem 2. ∎

It remains to pass to binary codes. For $q = 2^{2s}$ take the binary linear $[n = q-1, n-2s, 3]$ Hamming code and consider its orthogonal code, i.e., the simplex code. For simplicity let us augment each vector in it with a zero coordinate. This results in a binary linear code S of length q, dimension $2s$

and distance $q/2$ in which every nonzero vector has Hamming weight $q/2$. Establish a linear bijection between $\mathbb{F}_q$ and S and for a vector $c \in \mathscr{C}$ replace every coordinate by its image. We obtain a linear binary code C_n of length $n = qN$ and minimum distance $d_n := qN\delta/2$. Note that pairwise distances in $\mathscr{C}$ change by a factor $q/2$ upon passing to C_n, and so vectors of weight d_n in C_n are obtained from vectors of weight d in $\mathscr{C}$ and only from them. Together with Theorem 2 this completes the proof of Theorem 1.

Remarks. (1) From the definition of $E_q(\delta)$ we see that the interval (δ_1, δ_2) for large q is arbitrarily close to $(0, 1)$. Hence the result of Theorem 1 is valid for all values of d_n/n between 0 and 1/2.

(2) There are many possible choices for the code S in the final step. For instance, one could take $S = \{e_i, 1 \leqslant i \leqslant q\}$, where e_i is a binary q-vector with $e_{ij} = \delta_{i,j}$, $j = 1, ..., q$. Then the distances in $\mathscr{C}$ are doubled, and the qualitative argument of the proof is preserved. This gives a sequence of nonlinear codes C_n.

(3) The rate of the code C_n equals $2Rs/q$, where for large N the value $R > 0$ is given in the main theorem of [6].

(4) Upper bounds on the *average* weight spectrum of $\mathscr{C}$ over the choice of $D \in Div_a^+(X)$ for maximal curves were obtained in [7].

REFERENCES

1. A. Ashikhmin, A. Barg, and S. Litsyn, Estimates of the distance distribution of codes and designs, *IEEE Trans. Inform. Theory* **47** (2001), 1050–1061.
2. G. Kalai and N. Linial, On the distance distribution of codes, *IEEE Trans. Inform. Theory* **41** (1995), 1467–1472.
3. S. Litsyn, New upper bounds on error exponents, *IEEE Trans. Inform. Theory* **45** (1999), 385–398.
4. M. Rosenbloom and M. Tsfasman, Multiplicative lattices in global fields, *Invent. Math.* **101** (1990), 687–696.
5. M. Tsfasman and S. Vlăduţ, "Algebraic-Geometric Codes," Kluwer Academic, Dordrecht, 1991.
6. S. Vlăduţ, An exhaustion bound for algebro-geometric "modular" codes, *Problemy Peredachi Informatsii* **23** (1987), 28–41.
7. S. Vlăduţ, Two remarks on the spectra of algebraic geometry codes, *in* "Arithmetic, Geometry and Coding Theory" (R. Pellikaan, M. Perret, and S. G. Vlăduţ, Eds.), pp. 253–261, de Gruyter, Berlin, 1996.