Astérisque

GILLES COHEN
GÉRARD ZÉMOR

Subset sums and coding theory

Astérisque, tome 258 (1999), p. 327-339
<http://www.numdam.org/item?id=AST_1999__258__327_0>

Astérisque **258**, 1999, p. 327–339

SUBSET SUMS AND CODING THEORY

by

Gérard Cohen & Gilles Zémor

Abstract. — We study some additive problems in the group $(\mathbb{Z}_{/2\mathbb{Z}})^r$. Our purpose is to show how those problems are closely related to coding theory. We present some relevant classical coding techniques and make use of them to obtain some original contributions.

1. Introduction

Let G denote the group $\mathbf{F}^r$ where $\mathbf{F} = \{0, 1\}$ stands for the additive group with two elements. Let S be a generating set of G. For any positive integer i, denote by S^i the set of sums of i distinct elements of S. Set $S^0 = \{0\}$ and for any set I of non-negative integers, let $S^I = \cup_{i \in I} S^i$. Let us denote by $\rho(S)$ the smallest integer t such that any element of G can be expressed as a sum of t or less elements of S, i.e. such that

$$G = S^{[0,t]}.$$

Let us denote by $d(S)$ the smallest integer i such that 0 can be expressed as a sum of i distinct elements of S, i.e. let $d(S) - 1$ be the largest t such that

$$0 \notin S^{[1,t]}.$$

We wish to focus on the following 'additive' problems.

Problem 1. — *For given r and t, find the smallest s such that $|S| \geq s$ implies $\rho(S) \leq t$.*

Problem 2. — *For given r and t, find the largest s such that $|S| \leq s$ implies $\rho(S) \geq t$.*

Problem 3. — *For given r and d, find the smallest s such that $|S| \geq s$ implies $d(S) \leq d$.*

1991 ***Mathematics Subject Classification.*** — 94B05, 94B75, 05C25, 05C50, 11P99.
Key words and phrases. — Coding theory, additive theory.

Those three problems can be expressed as problems in coding theory. Indeed, problems 2 and 3 are classical coding problems of which we shall give a short self-contained presentation for the non specialist. Problem 1, although less known to coding theorists, is also amenable to coding techniques, and we shall present original contributions to it and also to the following generalisation of problem 3.

Problem 4. — *Given r and an arbitrary set of integers I, find the smallest s such that $|S| \geq s$ implies $0 \in S^I$.*

2. Coding-theoretic formulation of problems 1-4

What coding theorists call a (binary) *linear code* of length n is simply a subspace of the vector space $\mathbf{F}^n$. Let S be a generating set of $\mathbf{F}^r$ with $|S| = n$. There is an important linear code $C(S)$ associated to S whose coding-theoretic properties reflect the additive properties of S. To obtain it let $s_1, \ldots, s_n$ be any ordering of its elements that we shall write as column vectors. Consider the $r \times n$ matrix $\mathbf{H} = [s_1 \ldots s_n]$ and the associated function

$$\begin{array}{rcl} \sigma : \mathbf{F}^n & \to & G = \mathbf{F}^r \\ \mathbf{x} = (x_1 \ldots x_n) & \mapsto & \sigma(\mathbf{x}) = \mathbf{H}\,{}^t\mathbf{x} \end{array}$$

Define $C(S)$ to be the set of vectors $\mathbf{x}$ of $\mathbf{F}^n$ such that $\sigma(\mathbf{x}) = 0$. When defining such a code $C(S)$ associated to a set S we shall usually not specify which ordering $s_1, \ldots, s_n$ we are choosing because the properties of $C(S)$ that interest us are independent of it. To help distinguish between the two structures $G = \mathbf{F}^r$ and $\mathbf{F}^n$, we shall use plain letters to denote elements of G and bold letters to denote vectors of $\mathbf{F}^n$: furthermore, since the vector space structure of $\mathbf{F}^n$ will be used rather more heavily than that of G, we shall systematically refer to elements of $\mathbf{F}^n$ as *vectors*. $C(S)$ (or simply C when there is little ambiguity) is a subspace of $\mathbf{F}^n$ of dimension $k = n - r$. Its elements are referred to as *codewords*. $\mathbf{H}$ is called a *parity-check matrix* of C, and for any vector $\mathbf{x} \in \mathbf{F}^n$, $\sigma(\mathbf{x})$ is called the *syndrome* of $\mathbf{x}$. Two vectors $\mathbf{x} = (x_1 \ldots x_n)$ and $\mathbf{y} = (y_1 \ldots y_n)$ of $\mathbf{F}^n$ are said to be *orthogonal* if

$$\sum_{i=1}^{n} x_i y_i = 0$$

where computations are performed in $\mathbf{F}$. If C is a linear code of $\mathbf{F}^n$ of dimension k, then the set $C^\perp$ of vectors orthogonal to C is a linear code of dimension $n - k$. Any matrix $\mathbf{H}$ whose rows are independent vectors orthogonal to C make up a parity-check matrix of C.

Remark. — *Not every code C need be a code $C(S)$ for some set S. This is because not every code has a parity-check matrix with distinct columns.*

Coding theorists regard $\mathbf{F}^n$ as a metric space, i.e. endowed with the *Hamming distance* $d(\cdot, \cdot)$:

$$\begin{array}{rcl} \mathbf{F}^n \times \mathbf{F}^n & \to & [0, n] \\ (\mathbf{x}, \mathbf{y}) & \mapsto & d(\mathbf{x}, \mathbf{y}) \end{array}$$

where $d(\mathbf{x}, \mathbf{y})$ is defined as the number of coordinates where $\mathbf{x}$ and $\mathbf{y}$ differ. The *minimum distance* $d(C)$ of a code C is the smallest distance between a pair of distinct codewords,

$$d(C) = \min_{\substack{\mathbf{x},\mathbf{y}\in C \\ \mathbf{x}\neq\mathbf{y}}} d(\mathbf{x}, \mathbf{y}).$$

Note that $d(C)$ is also the minimum distance $d(\mathbf{x}, \mathbf{0})$ between the $\mathbf{0}$ vector and any non-zero codeword $\mathbf{x}$: this is because $d(\cdot, \cdot)$ is invariant by translation and C is an additive subgroup. The integer $d(\mathbf{x}, \mathbf{0})$ is called the *weight* of $\mathbf{x}$ and denoted by $w(\mathbf{x})$. The classical parameters of a linear code C are usually denoted by $[n, k, d]$ and refer respectively to its length, dimension and minimum distance.

Another classical parameter of a code C is its *covering radius* $\rho(C)$: it is the maximum distance between a vector of $\mathbf{F}^n$ and the code C, i.e.

$$\rho(C) = \max_{\mathbf{x}\in\mathbf{F}^n} d(\mathbf{x}, C)$$

where $d(\mathbf{x}, C) = \min_{\mathbf{c}\in C} d(\mathbf{x}, \mathbf{c})$.

Given a vector $\mathbf{x} = (x_1 \dots x_n)$ of $\mathbf{F}^n$, it is common to define its *support* by $supp(\mathbf{x}) = \{i, x_i = 1\}$. The syndrome of $\mathbf{x}$ can therefore be written as

$$\sigma(\mathbf{x}) = \sum_{i\in supp(\mathbf{x})} s_i$$

where the sum is computed in $\mathbf{F}^r$. It is now clear that the minimum distance of C equals the minimum cardinality of a subset I of S such that $\sum_{i\in I} s_i = 0$. In particular we have :

Remark. — *For any code C, there exists a set S not containing 0 such that $C = C(S)$ if and only if $d(C) \geq 3$.*

Similarly, it is readily checked that the covering radius of C is the smallest number of additions necessary to generate every non-zero element of $\mathbf{F}^r$ with elements of S. Summarizing,

Proposition 2.1. — *The correspondence $S \to C(S)$ is such that*

$$\begin{aligned} d(S) &= d(C(S)) \\ \rho(S) &= \rho(C(S)). \end{aligned}$$

The above correspondence transforms problems of an additive nature into *packing* and *covering* problems in a metric space. In particular, we see that problem 3 is equivalent to the fundamental problem of coding theory, namely determine the largest possible minimum distance of a linear code of length n and dimension k. There are several classical bounds relating n, k and d. Let us mention two simple bounds that we shall make use of later on.

***Proposition 2.2* (Hamming bound).** — *Any $[n, n-r, d]$ code satisfies*

$$\sum_{i=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{i} \leq 2^r.$$

Proof. — Since any vector $\mathbf{x} \in \mathbf{F}^n$ of weight $\leq d-1$ satisfies $\sigma(\mathbf{x}) \neq 0$, then all vectors with weight at most $\lfloor (d-1)/2 \rfloor$ must have distinct syndromes.

Using classical estimates for binomial coefficients, the Hamming bound states, asymptotically, that any $[n, nR, n\delta]$ code satisfies

$$R \leq 1 - h(\delta/2) + o(1) \tag{1}$$

where $h(x) = -x \log_2 x - (1-x)\log_2(1-x)$ denotes the binary entropy function.

***Proposition 2.3* (Varshamov-Gilbert bound).** — *Let n and r be given. There exists an $[n, n-r, d]$ code whenever*

$$\sum_{i=0}^{d-1} \binom{n-1}{i} < 2^r.$$

Proof. — We construct inductively a parity-check matrix of such a code. Suppose constructed an $r \times i$ matrix $\mathbf{H}_i$ such that any $d-1$ columns are linearly independent. They are at most N_i distinct linear combinations of columns involving at most $d-2$ terms, with

$$N_i = \sum_{j=1}^{d-2} \binom{n}{j}.$$

If $N_i < 2^r - 1$, then a nonzero element of $G = \mathbf{F}^r$ can be added to the set of columns of $\mathbf{H}_i$ to yield an $r \times (i+1)$ matrix $\mathbf{H}_{i+1}$ with the property that any $d-1$ of its columns are linearly independent ; equivalently $\mathbf{H}_{i+1}$ is the parity-check matrix of a code of minimal distance $\geq d$.

Asymptotically, the Varshamov-Gilbert bound reads: there exist $[n, nR, n\delta]$ codes with

$$R \geq 1 - h(\delta) + o(1). \tag{2}$$

There is no known better asymptotic lower bound on $R = k/n$. Let us just mention the most powerful upper bound on R due to McEliece, Rodemich, Rumsey, and Welch (see e.g. [**10**]) for a proof):

Proposition 2.4. — *Any $[n, nR, n\delta]$ code satisfies*

$$R \leq h\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right) + o(1). \tag{3}$$

Note that the Varshamov-Gilbert bound is not really constructive (the complexity of constructing a parity-check matrix for such codes is exponential in the length n). There are no known constructions of codes achieving the Varshamov-Gilbert bound for growing n and fixed R, $0 < R < 1$. There are, however, good constructions of codes with fixed d and growing n. We give a very short presentation of such codes, to which we shall refer later on.

Cyclic and BCH codes. — The one-to-one mapping

$$\mathbf{v} = (v_0, v_1, \ldots, v_{n-1}) \leftrightarrow v(X) = v_0 + v_1 X + \cdots + v_{n-1} X^{n-1}$$

gives us an identification of the binary vector space $\mathbf{F}^n$ with the additive structure of the algebra $\mathcal{A} = \mathbf{F}[X]/(X^n - 1)$. If a subspace of $\mathcal{A}$ has the additional property of being an ideal of the ring $\mathcal{A}$, it is called a *cyclic* code. Every ideal of $\mathcal{A}$ is principal and generated by a polynomial $g(X)$ which divides $X^n - 1$. Take n of the form $n = 2^m - 1$ so that $g(X)$ can be considered to have all its roots in the finite field $\mathbf{F}_{2^m}$ on 2^m elements. Let α be a primitive element of $\mathbf{F}_{2^m}$ and define the cyclic code C_e as the set of polynomials (modulo $X^n - 1$) whose roots contain $\alpha, \alpha^3, \ldots, \alpha^{2e-1}$ (since all these elements are roots of $X^n - 1$, this definition makes sense). It is a vector space over $\mathbf{F}$ of dimension at least $n - em$. Since these polynomials have their coefficients in $\mathbf{F}$, the set of their roots must be stable by the Frobenius homomorphism $x \mapsto x^2$, so that polynomials of C_e also have $\alpha^2, \alpha^4, \ldots, \alpha^{2e}$ as roots. Note that C_e can also be described as those vectors $\mathbf{v} = (v_0, \ldots, v_{n-1}) \in \mathbf{F}^n$ that are orthogonal to the rows of the matrix

$$\mathbf{H} = \begin{bmatrix} 1 & \alpha & \alpha^2 & \ldots & \alpha^{2^m-2} \\ 1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2(2^m-2)} \\ 1 & \alpha^3 & \alpha^6 & \ldots & \alpha^{3(2^m-2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha^{2e} & \alpha^{4e} & \ldots & \alpha^{2e(2^m-2)} \end{bmatrix}.$$

Now any $2e \times 2e$ submatrix of $\mathbf{H}$ is a van der Monde matrix and hence full-rank. Therefore any vector of $\mathbf{F}^n$ that is orthogonal to all the rows of $\mathbf{H}$ must have weight not less than $2e + 1$. The set C_e is called a BCH code: we have just proved that its parameters are

$$[n = 2^m - 1, k \geq 2^m - 1 - em, d \geq 2e + 1].$$

Cyclic and BCH codes have been extensively studied: see e.g. **[10]** and references therein. For fixed $d = 2e + 1$ and growing n they are, except for sporadic counter-examples, the best known constructions. Their dimension $k \geq n - e \log_2 n$ meets the asymptotic Hamming bound $k \leq n - e \log_2 n + 0(1)$.

Problem 2 is also classical, and can be reworded as: 'determine the smallest possible covering radius of a linear code of length n and dimension k'. We do not wish to dwell further on those two classical problems but rather refer to **[10]** for problem 3 and general background on coding theory, and to **[9]** for problem 2. Problem 4 is a generalisation of problem 3 that we shall discuss in section 4.

In the next section, we focus on problem 1 which is a truly additive problem in the sense that we are asking for those sets S, and their cardinalities, such that $S^{[1,t]} = S \cup S^2 \cup \cdots \cup S^t$ grows as slowly as possible.

3. Problem 1

Denote by $s_G(t)$ the smallest integer s such that, for any generating set S of $G = \mathbf{F}^r$, $\rho(S) \leq t$ whenever $|S| \geq s$. In other words $s_G(t) - 1$ is the largest cardinality of a

generating set S of G such that $\rho(S) > t$. Problem 1 asks for the determination of $s_G(t)$. Proposition 2.1 tells us that this can be understood as asking for the largest possible covering radius of a linear code $C(S)$ of given length. Because $\rho(S) = \rho(S \setminus \{0\})$, and in view of the remark preceding proposition 2.1, we are really asking for the largest covering radius of a linear code of given length and minimum distance $d \geq 3$.

3.1. A lower bound on $s_G(t)$. — It is natural to consider the following sets.

Definition 3.1. — *Call a τ-*cylinder *of $\mathbf{F}^r$, a subset isomorphic to $S = B_{\tau,1}(0) \times \mathbf{F}^{r-\tau}$, where $B_{\tau,1}(0)$ denotes the ball centered on 0 and of radius 1 in $\mathbf{F}^\tau$. In other words, for some properly chosen basis of $\mathbf{F}^r$, S is the subset of vectors of $\mathbf{F}^r$ whose first τ coordinates make up a vector of weight at most one. If S is a τ-cylinder of $\mathbf{F}^r$, then $\rho(S) = \tau$ and $|S| = (\tau + 1)2^{r-\tau}$.*

Since $s_G(t)$ must be larger than the cardinality of a $(t+1)$-cylinder, we have:

Proposition 3.1. — *Let $\log_2 |G| \geq t + 1$. Whenever $\log_2 |G| \geq t + 1$,*

$$s_G(t) > \frac{t+2}{2^{t+1}}|G|.$$

3.2. Upperbounding $s_G(t)$. — It is possible to prove that the above lower bound is the best possible for some values of r by a coding argument. The idea is to say, broadly speaking, that a code can't have too large a covering radius, otherwise, without changing the minimum distance, one would use it to construct a code with an impossibly large dimension.

Denote by $k(n, d)$ the maximum dimension of a linear code of length n and minimum distance at least d. We have, ([**9**])

Proposition 3.2. — *Let C be an $[n, k, d]$ linear code, and let ρ be its covering radius. We have:*

$$k + k(\rho, d) \leq k(n, d).$$

Proof. — Let $\mathbf{z}$ be a vector such that $d(\mathbf{z}, C) = \rho$ and of weight ρ. Assume, without loss of generality that its support is $supp(\mathbf{z}) = \{1, 2, \ldots, \rho\}$. Let C' be a code of length ρ and dimension $k(\rho, d)$. Let $(C'|\mathbf{0})$ be the code of length n obtained from C' by appending $\mathbf{0} \in \mathbf{F}^{n-\rho}$ to all words in C'. It is not difficult to check that the sum $C + (C'|\mathbf{0})$ is a code with minimal distance at least d and dimension $k + k(\rho, d)$.

One has, besides:

Lemma 3.1. — $k(n, 3) = n - 1 - \lfloor \log_2 n \rfloor$.

Proof. — To prove this, one just needs to find the smallest r such that there exists an $r \times n$ matrix with distinct non-zero columns.

Applying Proposition 3.2 and Lemma 3.1 we obtain the following upper bound on $s_G(t)$.

Proposition 3.3. — $s_G(t) \leq |G|/2^{t - \lfloor \log_2(t+1) \rfloor} + 1$.

Remarkably, the two bounds 3.1 and 3.3 coincide for t of the form $2^m - 2$, so that we have:

Corollary. — *For $m \geq 2$, $\log_2 |G| \geq 2^m - 1$, the following equality holds.*

$$s_G(2^m - 2) = |G|/2^{2^m - m - 1} + 1.$$

For the remaining values of t the upper and lower bounds of Propositions 3.1 and 3.3 leave a gap. For $t = 3$, the first value for which some uncertainty remains, we can obtain an improvement over Proposition 3.3. We make use of a theorem proved in [**13**] with a more traditional "additive" approach. It says that the subsets S of $\mathbf{F}^r$ such that $|S + S|$ is "small" tend to cluster around subgroups: the result is sharper than what can be said for general abelian groups. The precise statement is :

Theorem 3.1. — *Let S be a subset of $G = \mathbf{F}^r$. Let k be a nonnegative integer. One of the following holds.*

i. *There is a subgroup $H \neq \{0\}$ of G such that*

$$|S + H| - |S| < |H| + k$$

ii. *For any subset T of G such that $k \leq |T|^2 - 2$ and $2 \leq |G| - |S + T|$ we have*

$$|S + T| \geq |S| + |T| + k$$

Applying Theorem 3.1 with $k = 0$ yields:

Proposition 3.4. — $s_G(3) \leq |G|/3 + 1$.

Proof. — We prove that if S generates $G = \mathbf{F}^r$ and $|S| > |G|/3$ then $S + S + S = G$. We argue as follows. First check the result by hand for $\log_2 |G| \leq 4$. Then proceed by induction. If S satisfies the above, then:

1. Either $|S + S| \geq 2|S|$, and then $|S| + |S + S| > |G|$ so that the pigeon-hole principle implies $S + S + S = G$.
2. Or $|S + S| < 2|S|$, in which case Theorem 3.1 implies the existence of a non-trivial subgroup H of G such that $|S + H| - |S| \leq |H| - 1$. Consider now the partition $S = \cup S_i$ induced by the partition of G into cosets modulo H. Expressing $S + S$ as a union of sums $S_i + S_j$, we obtain by repeated application of the pigeon-hole principle that $S + S = S + S + H$. We finish by applying the induction hypothesis in G/H to the set of those cosets modulo H that intersect S, so as to obtain that $S + S + S$ intersects all cosets modulo H of G.

4. Constrained distances

In this section we consider problem 4, i.e. studying large sets S such that $0 \notin S^I$ for arbitrary $I \subset [1, |S|]$. Let us restate the problem in coding terms. We shall use the notation of [**5**]. For a code C, let

$$D(C) = \{w(\mathbf{c}) \mid \mathbf{c} \in C, \mathbf{c} \neq \mathbf{0}\}$$

$$l(n, D) = \max\{\dim C \mid D(C) \subset D\}.$$

Denote by $\overline{D} = [1, n] \setminus D$ the complement of D.

If $D \subset D(C)$, C is sometimes called a *D-clique*. The classical coding case is $D = [d, n]$, but the function $l(n, D)$ can vary very much with the nature of the set

D. For instance D-cliques with $D = [0, d]$, in other words sets with *maximal distance* d, have been considered under the name of *anticodes* [**6**]. These anticodes have been used to construct good codes, see ch. 17 §6 of [**10**]. More recently the problem of forbidding one distance, i.e. studying $l(n, \{\overline{d}\})$, has been considered. A variety of approaches to the problem have been put forward, among which additive techniques and more traditional coding approaches. By way of illustration, let us mention the problem of determining $l(4t, \{\overline{2t}\})$. It was conjectured by Ito that $l(4t, \{\overline{2t}\}) = 2t$. An elegant proof was found by Alon in the case $4t = 2^m$, using the following theorem of Olson. For an abelian group G, denote by $s(G)$ the smallest positive integer such than any sequence $g_1 \dots g_s$ of (not necessarily distinct) non-zero elements of G contains a subsequence summing to zero. Olson's Theorem [**12**] states.

Theorem 4.1. — *Consider the finite abelian p-group* $G = \mathbb{Z}/p^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p^{e_k}\mathbb{Z}$ *; then* $s(G) = 1 + \sum_{i=1}^{k}(p^{e_i} - 1)$.

Sketch of Alon's proof. — Let C be a linear code of length $4t$ and dimension $2t + 1$. Consider the columns of a $(2t - 1) \times 4t$ parity-check matrix of C and add to each of them an extra coordinate consisting of the 1 element of the group $\mathbb{Z}/2^{m+1}\mathbb{Z}$. Thus we are dealing with $4t$ elements of the group $G = (\mathbb{Z}/2\mathbb{Z})^{2t-1} \times \mathbb{Z}/2^{m+1}\mathbb{Z}$. Olson's Theorem implies $s(G) = 4t - 1$, hence the existence of a proper subset of those elements that sum to zero: because of the last coordinate this subset must consist of exactly $2t$ elements and therefore correspond to a word of C of weight $2t$.

Ito's conjecture was finally proved in [**5**] for all t.

4.1. General results. — Most of the results of this section carry over to non linear codes: we shall not concern ourselves with these generalisations, however, since they would take us too far from our additive motivation.

Let us start by a general result of Delsarte [**4**].

Theorem 4.2

$$2^{l(n,D)} \leq \sum_{i=0}^{|D|} \binom{n}{i}.$$

We present a concise proof of this classical result which should give some flavour of the methods of coding theory.

Proof of Theorem 4.2. — Let C be a code with parity-check matrix $\mathbf{H}$. Let S be the set of the columns of $\mathbf{H}$. Let us associate to C the Cayley graph $\mathcal{C}$ defined as having $G = \mathbf{F}^r$ as vertex set, and edge set $\{(g, g + s) \mid g \in G, s \in S\}$. Let $\mathbf{A} = (a_{uv})$ be the adjacency matrix of $\mathcal{C}$, i.e. the matrix whose rows and columns are indexed by G and such that

$$a_{uv} = \begin{cases} 1 \text{ if } v = u + s, \ \ s \in S \\ 0 \text{ otherwise} \end{cases}$$

Notice that the quantity $\rho(C) = \rho(S)$ is exactly the diameter Δ of $\mathcal{C}$.

Let $Spe(\mathcal{C})$ be the set of eigenvalues of $\mathbf{A}$. The following lemma is classical in graph theory.

Lemma 4.1. — *Suppose the graph $\mathcal{C}$ has diameter Δ. Then*

$$\Delta + 1 \leq |Spe(\mathcal{C})|.$$

Proof of lemma 4.1. — Recall that the entry in position (u, v) of the matrix $\mathbf{A}^i$ equals the number of walks $u = u_0, u_1 \ldots, u_i = v$ of length i from vertex u to vertex v. Consider the algebra $\mathbf{C}[\mathbf{A}]$ of polynomials in $\mathbf{A}$. On the one hand, it is standard linear algebra that $\dim \mathbf{C}[\mathbf{A}] = |Spe(\mathcal{C})|$. On the other hand, whenever $i \leq \Delta$, there is a walk of length i from some vertex u to some vertex v such that no walk of length $< i$ exists between u and v. This means that $(\mathbf{A}^i)_{uv} \neq 0$ while $(\mathbf{A}^j)_{uv} = 0$ for $j < i$. Therefore $\{\mathbf{I}, \mathbf{A}, \ldots, \mathbf{A}^\Delta\}$ is a linearly independent set in $\mathbf{C}[\mathbf{A}]$.

Now in our particular case, it is straightforward to check that for any character χ of G, $[\chi(v)]_{v \in G}$ is an eigenvector of $\mathbf{A}$ associated to the eigenvalue

$$\lambda_\chi = \sum_{s \in S} \chi(s).$$

Every character of the group $G = \mathbf{F}^r$ is of the form

$$\chi_u : v \mapsto (-1)^{(u|v)}$$

for some $u \in G$, where $(u|v)$ denotes the scalar product in $\mathbf{F}^r$. So we see that $\lambda_{\chi_u} = n - 2w({}^t u.\mathbf{H})$, so that the number of distinct eigenvalues of $\mathcal{C}$ is exactly the number of distinct weights in the subspace generated by the rows of $\mathbf{H}$, i.e. the dual code $C^\perp$ of C. Summarizing, we have:

***Theorem 4.3* (Delsarte).** — $\rho(C) \leq |D(C^\perp)|$.

Note now that, from the definition of $\rho(S) = \rho(C)$, one has the inequality

$$\sum_{i=0}^{\rho(C)} \binom{n}{i} \geq 2^r. \tag{4}$$

Relation (4) together with Theorem 4.3 prove Theorem 4.2.

Let us state the following result from [**5**].

Proposition 4.1. — *For $n \geq 4t$,*

$$\begin{aligned} l(n, \{\overline{2t}\}) &\leq n - 2t \\ l(n, \{\overline{2t, 2t+1}\}) &\leq n - 2t - 1. \end{aligned}$$

We shall now derive a variation on the so-called "Elias-Bassalygo lemma" [**1**].

Denote by $A(n, D)$ the maximal size of a (not necessarily linear) subset of $\mathbf{F}^n$ such that any two of its elements have distance in D.

Denote by $A(n, D, w)$ the maximal size of a subset of $\mathbf{F}^n$ such that any two of its elements have weight w and distance in D.

Proposition 4.2

$$A(n,D) \leq \frac{2^n}{\binom{n}{w}} A(n,D,w).$$

Proof. — Let C be a code (simply a set of vectors in the non-linear case) realizing $A(n,D)$. Consider its 2^n translates $C+\tau, \tau \in \mathbf{F}^n$. Each vector of $\mathbf{F}^n$, and in particular those of weight w, appear $A(n,D)$ times in the union of the translates $C+\tau$. Thus one of the translates, in itself a D-clique because $d(\cdot,\cdot)$ is invariant by translation, must contain at least $\binom{n}{w} A(n,D) 2^{-n}$ vectors of weigh w. Hence

$$\binom{n}{w} A(n,D) 2^{-n} \leq A(n,D,w).$$

4.2. Forbidding one distance. — We shall need the following result [**7**].

Proposition 4.3. — *If $\mathcal{F}$ is a family of w-subsets of an n-set no two of which intersect in exactly e elements, then*

$$|\mathcal{F}| \leq c_w n^{\max\{e, w-e-1\}}$$

where c_w is a constant depending only on w.

Set $w = d = 2e$, then clearly any two members of a family achieving $A(n, \overline{2e}, 2e)$ do not intersect in e elements. Thus Proposition 4.3 yields

$$A(n, \overline{2e}, 2e) \leq c_{2e} n^e$$

and by Proposition 4.2 we get, fixing e and letting n go to infinity,

$$A(n, \overline{2e}) = O\left(\frac{2^n}{n^e}\right).$$

Hence,

$$l(n, \overline{2e}) \leq n - e \log_2 n + O(1). \tag{5}$$

In other words, for fixed e, it is asymptotically just as costly to forbid the distance $2e$ between codewords as to forbid all distances d, $1 \leq d \leq 2e$, since BCH codes meet (5). We have:

Proposition 4.4. — $l(n, \overline{2e}) = n - e \log_2 n + O(1)$.

We now consider the case when the forbidden distance d increases linearly with n. In other words, we fix λ and study $l(n, \overline{\lambda n})$ by which we mean $l(n, \overline{\{\lfloor \lambda n \rfloor\}})$. Some caution is in order when dealing with the asymptotical behaviour of $n^{-1} l(n, \overline{\lambda n})$, since this function of n does not converge: indeed, the $[n, n-1, 2]$ even weight subcode of $\mathbf{F}^n$ shows that $l(n, \overline{2e+1}) = n-1$, hence $\limsup n^{-1} l(n, \overline{\lambda n}) = 1$. We suspect that the sequence $n^{-1} l(n, \overline{\lambda n})$ actually has many accumulation points. We shall now derive a result on $\liminf n^{-1} l(n, \overline{\lambda n})$.

We shall need the following from [**8**]:

Proposition 4.5. — *Let q be a prime power. Let $\mathcal{F}$ be a set of w-subsets of the n-set $\{1, 2, \ldots, n\}$. Suppose that for any distinct $F, F' \in \mathcal{F}$ we have*

$$|F \cap F'| \not\equiv w \bmod q$$

then

$$|\mathcal{F}| \leq \binom{n}{q-1}.$$

We now obtain:

Proposition 4.6

$$\liminf_{n\to\infty} n^{-1} l(n, \overline{\lambda n}) \leq 1 - h(\lambda) + h(\lambda/2) + o(1).$$

Proof. — Suppose d equals twice the power of a prime $d = 2q$. Let $w = 2q - 1$. Any code of constant weight w and such that no two codewords are at distance d from each other yields a set $\mathcal{F}$ such that $|F \cap F'| \not\equiv -1 \bmod q$ for distinct $F, F' \in \mathcal{F}$. Hence

$$A(n, \overline{2q}, 2q-1) \leq \binom{n}{q-1} \leq 2^{n(h(\lambda/2)+o(1))}.$$

Apply Proposition 4.2 to conclude the proof.

Note that for $\lambda < 0.27$, this improves on Proposition 4.1.

4.3. Forbidding multiples of a given distance. — More generally, if q is a prime power and $\lambda n = 2iq$, considering constant weight codes of weight $w = (i+1)q - 1$, one obtains

Proposition 4.7

$$n^{-1} l(n, \overline{\{2q, 4q, \ldots, 2iq\}}) \leq 1 - h\left(\frac{i+1}{2i}\lambda\right) + h\left(\frac{\lambda}{2i}\right) + o(1).$$

Remark. — *For growing i, the right hand side of this last inequality tends to $1 - h(\lambda/2)$, so that it can be considered as a refinement of the Hamming bound (1)*

$$n^{-1} l(n, \overline{[1, \ldots, \lambda n]}) \leq 1 - h(\lambda/2)$$

in the sense that one need not forbid every distance in $[1, \ldots, \lambda n]$.

4.4. A construction. — We have the lower bound:

Proposition 4.8. — *For $\lambda \leq 1/3$,*

$$n^{-1} l(n, \overline{\lambda n}) \geq 1 - (1-\lambda) h\left(\frac{\lambda}{1-\lambda}\right) + o(1).$$

Proof. — Consider the generating matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{\lambda n-1} & 0 \\ 0 & \mathbf{G}_0 \end{bmatrix}$$

where $\mathbf{G}_0$ is a generator matrix of an optimal code C_0 of length $n - \lambda n + 1$ and distance $\lambda n + 1$. Obviously every combination of rows of $\mathbf{G}$ has weight at most $\lambda n - 1$ - if it does not use rows of $\mathbf{G}_0$ - or at least $\lambda n + 1$ if it does.

Take for C_0 a code lying on the Varshamov-Gilbert bound (2) to get the asymptotical result.

Large gaps remain between upper and lower bounds.

Open problem. — *It would be particularly interesting to known what is the most "persistent" distance in linear codes, in other words, what is the value of λ that minimizes* $\liminf n^{-1} l(n, \overline{\lambda n})$ *?*

5. Intersecting codes

We would like to conclude by another intriguing problem with an additive flavour. Let us say that a subset $S = \{s_1, \ldots, s_n\}$ of an abelian group G has the *intersecting property* if there do not exist two disjoint subsets I and J of $[1, n]$ such that both

$$\sum_{i \in I} s_i = 0 \qquad \text{and} \qquad \sum_{j \in J} s_j = 0.$$

An *intersecting code* C is a linear code with the property that any two non-zero codewords have intersecting supports. Equivalently, it is a code C such that the set of columns of any parity-check matrix of C has the intersecting property in $G = \mathbf{F}^r$.

Problem 5. — *Given r, what is the maximal size $\iota(r)$ of $S \subset \mathbf{F}^r$ with the intersecting property ?*

This problem was first investigated by Miklós [**11**], and has since proved to lead to a variety of applications, see [**2**]. A lower bound on $\iota(r)$ can be derived by random arguments [11,2] Asymptotically it reads:

$$\iota(r) \geq \frac{2r}{\log_2 3} \approx 1.26r.$$

To obtain an upper bound, notice that an intersecting code must have $d \geq k$. Otherwise choose a minimum weight codeword $\mathbf{c}$: among the 2^k codewords there must be two, $\mathbf{c}'$ and $\mathbf{c}''$, that coincide on the d coordinates of the support of $\mathbf{c}$. Therefore $\mathbf{c}$ and $\mathbf{c}' + \mathbf{c}''$ have nonintersecting supports, a contradiction. This argument, namely $\delta \geq R$, together with the bound (3) gives

$$\iota(r) \leq 1.40r.$$

References

[1] Bassalygo L.A., *New bounds for error-correcting codes.*, Problemy Peredachi Informatsii, **1**, 1965, 41–45.

[2] Cohen G. and Zémor G., *Intersecting codes and independent families*, IEEE Trans. on Inf. Theory, **40**, 1994, 1872–1881.

[3] Cohen G.D., Karpovsky M., Mattson H.F. Jr. and Schatz J., *Covering radius - survey and recent results*, IEEE Trans. Inf. Theory, **31**, 1985, 328–344.

[4] Delsarte P., *Four fundamental parameters of a code and their combinatorial significance*, Info. and control, **23**, 1973, 407–438.

[5] Enomoto H., Frankl P., Ito N. and Nomura K., *Codes with given distances*, Graphs and Combinatorics, **3**, 1987, 25–38.

[6] Farrell P.G., *Linear binary anticodes*, Electronics Letters, **6**, 1970, 419–421.

[7] Frankl P. and Füredi Z., *Forbidding just one intersection*, J.C.T. A, **39**, 1985, 160–176.

[8] Frankl P. and Wilson R.M., *Intersection theorems with geometric consequences*, Combinatorica, **1**, 1981, 357–368.

[9] Godlewski P., *WOM-codes construits à partir des codes de Hamming*, Discrete Math., **65**, 1987, 237–243.

[10] MacWilliams F.J. and Sloane N.J.A., *The theory of error-correcting codes*, North-Holland, 1977.

[11] Miklós D., *Linear binary codes with intersection properties*, Discrete Applied Math., **9**, 1984, 187–196.

[12] Olson J.E., *A combinatorial problem on finite abelian groups*, J. Number Theory, **1**, 1969, 195–199.

[13] Zémor G., *Subset sums in binary spaces*, Europ. J. of Combinatorics, **13**, 1992, 221–230.

GÉRARD COHEN, Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, 75 634 Paris Cedex 13, France • *E-mail :* `cohen@inf.enst.fr`

GILLES ZÉMOR, Ecole Nationale Supérieure des Télécommunications, 46 rue Barrault, 75 634 Paris Cedex 13, France • *E-mail :* `zemor@res.enst.fr`