$(2,1)$-separating systems beyond the probabilistic bound

Hugues Randriambololona

January 10, 2012

1 Introduction

One of the most powerful tools to derive lower bounds in extremal combinatorics is the so called *probabilistic method* [1]. Roughly speaking, to prove the existence of an object of a given size satisfying certain conditions, one shows that a random object of this size (maybe after being slightly modified) has a positive probability to satisfy these conditions.

In many problems the lower bound given by this method is conjectured exact, at least asymptotically, and sometimes one can prove it is indeed so. This means that optimal solutions to such problems are rather common. On the other hand, when the probabilistic lower bound is not asymptotically exact, optimal solutions tend to be rare and have some particular structure. So, from a theoretical point of view, it is of great importance to know whether a problem belongs to one or the other of these two classes.

The problem we will be dealing with in this paper is that of $(2,1)$-separation. As can be seen from [22], this problem, and more generally the theory of separating systems, has a quite long history. While its origins could be arguably traced back to [20], its first appearance, in the precise form we will be interested in, can be found in [7], motivated by a problem in electrical engineering. In fact the notion of separation there defined is very natural and ubiquitous, and several authors have introduced and studied equivalent versions, sometimes independently, in various contexts and in various languages (e.g. frameproof codes, intersecting codes, covering arrays, hash families...). We point out the following two elegant formulations which can be found in [11], the first being in terms of information theory (dealing with binary sequences), the second in terms of extremal combinatorics (dealing with set systems):

Problem A. *How many different points can one find in the n-dimensional binary Hamming space so that no three of them are on a line?*

(We say three points in a metric space are on a line if they satisfy the triangle inequality with equality.)

Problem A*. *How many different subsets can one find in an n-set so that no three A, B, C of them satisfy $A \cap B \subset C \subset A \cup B$?*

The equivalence between these two formulations is seen by identifying each binary sequence with its support set.

We will take the condition in Problem A as the definition of a $(2,1)$-separating system. Of course it can be extended to larger alphabets. We will only use the case where the alphabet is a (finite) field K:

Definition 0. *A $(2,1)$-separating code over K of length n is a subset $C \subset K^n$ such that any pairwise distinct $x, y, z \in C$ satisfy*

$$d(x,z) < d(x,y) + d(y,z) \tag{1}$$

where d is the Hamming distance in K^n.

Condition (1) can be rephrased as saying that there is a coordinate i such that $y_i \notin \{x_i, z_i\}$, or equivalently, $(x_i - y_i)(z_i - y_i) \neq 0$ in K. If moreover C is linear, this says in turn that any two non-zero codewords have intersecting supports. Such a code is then called a *linear intersecting code* [14][15][4].

There are higher notions of (s,t)-separating codes, and several other variants; see the survey [22] (of notable interest in the literature, one also finds the terminology s-frameproof for $(s,1)$-separating codes, and s-secure-frameproof for (s,s)-separating codes [3][24]). These can be defined either in terms of coordinates, or in terms of metric convexity and the Hahn-Banach property, generalizing (1). For a further discussion of these ideas, and the analogy between Hamming and Euclidean spaces in this regard, see [18]. We will not need this material here, and focus on our main topic which is as follows.

Denote by $M(n)$ the common solution to Problems A and A*, and define its asymptotic exponent

$$\rho = \limsup_{n\to\infty} \frac{\log_2 M(n)}{n}. \tag{2}$$

It is shown in [11] that ρ satisfies the inequalities

$$1 - \frac{1}{2}\log_2 3 \leq \rho \leq \frac{1}{2} \tag{3}$$

where the derivation of the lower bound

$$1 - \frac{1}{2}\log_2 3 \approx 0.207518\ldots \tag{4}$$

is a typical example of use of the probabilistic method (it also follows from the earlier works [21][10][15][4], some of them of a more coding-theoretic nature, but still non-constructive).

The reader certainly noticed there is plenty of space for improvement between the two bounds in (3), and indeed the main aim of this paper will be to reduce this gap, although by an admittedly modest quantity:

Theorem 1. *The asymptotic exponent ρ satisfies the lower bound*

$$\rho \geq \frac{3}{50}\log_2 11 \approx 0.207565\ldots \tag{5}$$

As we will see, the proof of this theorem is fully constructive. And, however small the improvement from (4) to (5) might be, it is positive enough to ensure this new construction stands beyond the probabilistic bound. In fact, from the author's viewpoint, the tininess of this improvement makes it even nicer, since it results from an almost miraculous numerical coincidence. Do Mathematics have a sense of humour?

Our construction improves on the one of [5], section 7.2, while using the same *concatenation* argument (indeed it is easily seen from (1) that concatenating two $(2,1)$-separating codes gives a $(2,1)$-separating code again). The codes to be concatenated are chosen as follows:

- The outer code is a linear intersecting (hence $(2,1)$-separating) algebraic geometry code over $\mathbb{F}_{121}$.
- The inner code is any subcode of size $M = 121$ out of the 128 codewords of the binary non-linear *one-shortened Nordstrom-Robinson code* of length $n = 15$. Remark the only possible distances in this code are 0, 6, 8, and 10, so it satisfies (1).

Expressing the rate of the concatenated code as the product of the inner and outer rates then gives the lower bound

$$\rho \geq \frac{\log_2 121}{15} R_{121} \tag{6}$$

where R_q denotes the asymptotic maximal achievable rate for linear intersecting codes over $\mathbb{F}_q$.

This choice of parameters, and especially the choice of $q = 121$, is the result of a certain trade-off that can only be justified a posteriori. Known lower bounds on R_q get better as q grows. But on the other side, as in the binary case, we have the upper bound $R_q \leq \frac{1}{2}$ (see e.g. [2]), and when q and the length of the inner code grow, the rate of the inner code becomes limited by the asymptotic exponent ρ. Thus there is no hope to get a construction of rate significantly better than $\frac{\rho}{2}$ by this concatenation argument if q is taken too large. In fact, candidates for the inner code have higher rate for smaller lengths, which implies to keep q of moderate size.

It turns out that the Nordstrom-Robinson code is a $(2,1)$-separating code with exceptionally high rate considering its length. It is the first of a sequence of Kerdock codes, that can be shown $(2,1)$-separating by the very same method [12], but whose parameters are of lesser interest for concatenation. Remark also that, to use the full power of algebraic geometry codes, we need q to be a square. By luck, our square 121 is quite close to 128, hence restricting the Nordstrom-Robinson code to a 121-subcode has only marginal impact on the rate.

In [5] the authors remark that a linear code of relative minimum distance larger than one-half is intersecting. Combined with the Tsfasman-Vladut-Zink

bound [26] this gives

$$R_q \geq \frac{1}{2} - \frac{1}{A(q)} \tag{7}$$

where $A(q) = q^{1/2} - 1$ if q is a square. Hence $R_{121} \geq 0.4$ and $\rho \geq 0.184503$ which was the best constructive lower bound up to now.

In [27] Xing gives a new criterion for an AG code to be intersecting, that does not rely on the minimum distance of the code. From this criterion and a (non-constructive) counting argument in the Jacobian of the curve he deduces

$$R_q \geq \frac{1}{2} - \frac{1}{A(q)} + \frac{1 - 2\log_q(2)}{2A(q)} \tag{8}$$

hence $R_{121} \geq 0.435546$ and $\rho \geq 0.200877$ which is still below (4). However we will improve on Xing's bound (8) as follows:

Theorem 2. *Let q be a prime power with $A(q) > 4$. Then the asymptotic maximal achievable rate for linear intersecting codes over $\mathbb{F}_q$ satisfies*

$$R_q \geq \frac{1}{2} - \frac{1}{2A(q)}. \tag{9}$$

Moreover if $q \geq 25$ is a square, then $R_q \geq \frac{1}{2} - \frac{1}{2(q^{1/2}-1)}$.

This new bound was first conjectured in [17] and, as noted there, it implies Theorem 1. Indeed, it gives $R_{121} \geq \frac{9}{20} = 0.45$, hence combined with (6):

$$\rho \geq \frac{\log_2 121}{15}\frac{9}{20} = \frac{3}{50}\log_2 11 > 0.207565. \tag{10}$$

Thus the rest of this paper will be devoted to the proof of Theorem 2. We will do so by giving an effectively constructible family of linear intersecting codes attaining (9) — at least, provided an effectively constructible family of curves attaining $A(q)$ is known, which will be true in our case of interest.

2 Algebraic geometry codes and the intersecting support property

Here we recall some material from [27], and start to develop from it.

Let K be a field (in the next section we will also suppose K perfect, and actually the reader may assume K is a finite field).

If X is an algebraic curve (a smooth, projective, absolutely irreducible 1-dimensional scheme) over K, of genus g, and D is a divisor on X (in this text "divisor" will always mean "K-rational divisor"; likewise, "points" will be "K-points", etc.) we denote by $\mathcal{L}(D) = \Gamma(X, \mathcal{O}(D))$ its space of global sections, and $l(D) = \dim_K \mathcal{L}(D)$ the dimension of the latter.

Recall X admits a so called *canonical divisor* Ω, which may be taken to be the divisor of any (rational) differential form on X. It has degree $\deg(\Omega) = 2g-2$ and dimension $l(\Omega) = g$. The Riemann-Roch theorem asserts that

$$l(D) = \deg(D) + 1 - g + l(\Omega - D). \tag{11}$$

In particular $l(D) \geq \deg(D) + 1 - g$, with equality when $\deg(D) \geq 2g - 1$.

Suppose given a divisor G on X that can be written as a sum of distinct (K-)points, each with multiplicity 1. Let $n = \deg(G) \leq |X(K)|$ be its degree, and choose an ordering $P_1, \ldots, P_n$ of the points in its support, so

$$G = P_1 + \cdots + P_n. \tag{12}$$

Also, for each i, choose a local parameter t_i at P_i. Then, if D is any divisor on X, the section $t_i^{-v_{P_i}(D)}$ is a trivialization for $\mathcal{O}(D)$ at P_i. Restricting to the fiber, this trivialization then gives an identification $\mathcal{O}(D)|_{P_i} \simeq K$. Now combining these identifications with the natural restriction map

$$\mathcal{L}(D) \longrightarrow \bigoplus_{i=1}^{n} \mathcal{O}(D)|_{P_i} \tag{13}$$

leads to the following:

Definition 3. *For any divisor D on X, the generalized Goppa evaluation code $C(G, D)$ is the image of the morphism*

$$\begin{array}{rccc} \phi_{G,D}: & \mathcal{L}(D) & \longrightarrow & K^n \\ & f & \mapsto & ((t_1^{v_1} f)(P_1), \ldots, (t_n^{v_n} f)(P_n)) \end{array} \tag{14}$$

where for each i we let $v_i = v_{P_i}(D)$.

The kernel of $\phi_{G,D}$ is $\mathcal{L}(D-G)$. Hence, if $l(D-G) = 0$, then $\dim C(G, D) = l(D)$. This occurs for example when $\deg(D) < n$.

As noted by Xing, this construction generalizes Goppa's evaluation codes, while allowing the supports of G and D to overlap (in fact Xing's original definition also asked D to be positive, but this condition is clearly unnecessary).

A virtue of this description is that the ordering of the P_i and the choice of the t_i are made once and for all, independently of D. This gives some coherence in the choice of our identifications of the fibers $\mathcal{O}(D)|_{P_i} \simeq K$ as D varies, which in turn makes the system of our evaluation maps $\phi_{G,D}$ compatible, in the sense that, given two divisors D and D', the following diagram is commutative:

$$\begin{array}{ccc} \mathcal{L}(D) \times \mathcal{L}(D') & \xrightarrow{\phi_{G,D} \times \phi_{G,D'}} & K^n \times K^n \\ \downarrow & & \downarrow \\ \mathcal{L}(D + D') & \xrightarrow{\phi_{G,D+D'}} & K^n \end{array} \tag{15}$$

where the first vertical map is multiplication in the function field $K(X)$, and the second vertical map is termwise multiplication in K^n. Indeed, both paths in the diagram send $(f, f') \in \mathcal{L}(D) \times \mathcal{L}(D')$ to $((t_1^{v_1+v_1'} ff')(P_1), \ldots, (t_n^{v_n+v_n'} ff')(P_n))$ in K^n.

Said otherwise, the collection of maps $\phi_{G,D}$ define a morphism of K-*algebras*

$$\phi_G : \bigoplus_{D \in \mathrm{Div}(X)} \mathcal{L}(D) \longrightarrow K^n \tag{16}$$

where the multiplication law in K^n is termwise multiplication.

We now recall:

Theorem 4 (Xing's criterion, [27] Th. 3.5, with $s = 2$)**.** *With the preceding notations, suppose* $\deg(D) < n$ *and*

$$l(2D - G) = 0. \tag{17}$$

Then $C(G, D)$ has dimension $l(D)$ and is a linear (self-)intersecting code.

In fact it is possible to say slightly more.

Two linear codes $C, C' \subset K^n$ are said mutually intersecting if any non-zero codewords $c \in C$ and $c' \in C'$ have intersecting supports. Then:

Proposition 5. *Suppose D, D' are divisors on X with* $\deg(D) < n$, $\deg(D') < n$, *and*

$$l(D + D' - G) = 0. \tag{18}$$

Then $C = C(G, D)$ and $C' = C(G, D')$ have dimension $l(D) \geq \deg(D) + 1 - g$ and $l(D') \geq \deg(D') + 1 - g$ respectively, and are mutually intersecting.

Proof. Let $c \in C$ and $c' \in C'$ and suppose the termwise product cc' is zero in K^n. Write $c = \phi_{G,D}(f)$ and $c' = \phi_{G,D'}(f')$ for $f \in \mathcal{L}(D)$ and $f' \in \mathcal{L}(D')$. Then $0 = cc' = \phi_{G,D+D'}(ff')$ so $ff' \in \ker \phi_{G,D+D'} = \mathcal{L}(D + D' - G) = \{0\}$, hence $f = 0$ or $f' = 0$, that is $c = 0$ or $c' = 0$. This proves the intersection property, and then the lower bound on the dimensions follows from Riemann-Roch. □

Remark that this proposition includes Theorem 4 as a particular case (namely when $D = D'$). In fact the proof given here is essentially the same as Xing's.

We will use this variant of Xing's criterion to give a lower bound on the rates of pairs of mutually intersecting codes. While easier, the proof of this result will serve as a model for the proof of Theorem 2 in the last section; and it is also certainly of independent interest.

Lemma 6. *Let X be a curve over K of genus g, and let A be a divisor on X with* $\deg(A) \leq g - 2$ *and*

$$l(A) = 0. \tag{19}$$

Then for all points $P \in X(K)$ except perhaps for at most g of them, we have

$$l(A + P) = 0. \tag{20}$$

Proof (adapted from [23], ch. I, claim (6.8)). By contradiction, suppose there are $g+1$ distinct points $P_1, \ldots, P_{g+1} \in X(K)$ for which (20) fails, hence for each $1 \leq i \leq g+1$ we can find a function $f_i \in \mathcal{L}(A+P_i) \setminus \mathcal{L}(A)$. Let also

$$A' = A + P_1 + \cdots + P_{g+1}. \tag{21}$$

Then we also have $f_i \in \mathcal{L}(A') \setminus \mathcal{L}(A' - P_i)$, which means that the quotient space $\mathcal{L}(A')/\mathcal{L}(A'-P_i)$ has dimension 1 and admits f_i as a generator. On the other hand, for $j \neq i$, we have $f_i \in \mathcal{L}(A'-P_j)$, hence f_i maps to 0 in $\mathcal{L}(A')/\mathcal{L}(A'-P_j)$. This implies that the natural map

$$\mathcal{L}(A') \longrightarrow \bigoplus_{i=1}^{g+1} \mathcal{L}(A')/\mathcal{L}(A'-P_i) \tag{22}$$

is onto, hence by the rank theorem,

$$l(A') \geq g+1. \tag{23}$$

But at the same time

$$\deg(A') = \deg(A) + g + 1 \leq 2g-1 \tag{24}$$

which contradicts Riemann-Roch. □

Proposition 7. *Suppose $n > g$. Let m be an integer with $g \leq m < n$ and let D be a divisor on X of degree $\deg(D) = m$. Then there exists a divisor D' on X, with $\deg(D') = n+g-1-m$, such that $C = C(G,D)$ and $C' = C(G,D')$ have dimension*

$$\dim C \geq m+1-g \tag{25}$$

and

$$\dim C' \geq n-m \tag{26}$$

respectively, and are mutually intersecting.

(Note the hypothesis $n > g$ implies $|X(K)| > g$, hence in some way "X has many rational points".)

Proof. For $0 \leq i \leq g$ we construct divisors D'_i such that

$$\deg(D'_i) = n+i-1-m \qquad \text{and} \qquad l(D+D'_i-G) = 0 \tag{27}$$

iteratively as follows:

- Start with any divisor D'_0 of degree $n-1-m$, hence $\deg(D+D'_0-G) < 0$ and $l(D+D'_0-G) = 0$ as asked.
- Suppose up to some $i < g$, we have a divisor D'_i satisfying (27). The divisor $A = D+D'_i-G$ has then degree $\deg(A) = i-1$ and $l(A) = 0$, and since $|X(K)| > g$, we can apply Lemma 6 to find P such that $l(A+P) = 0$. Then we set $D'_{i+1} = D'_i + P$, so D'_{i+1} satisfies (27).

- This ends when $i = g$, and we set $D' = D'_g$.

With this choice of D', the conditions in Proposition 5 are satisfied, hence C and C' are mutually intersecting. Moreover, $\dim C' = l(D') \geq \deg(D') + 1 - g = n - m$ as claimed. □

Let q be a prime power. Say that a sequence of curves X_i over the finite field $\mathbb{F}_q$ form an ∞-sequence if the genus g_i of X_i tends to infinity as i goes to infinity.

Let $A(q)$ be the *largest* real number such that there exists an ∞-sequence of curves X_i over $\mathbb{F}_q$ with

$$\frac{|X_i(\mathbb{F}_q)|}{g_i} \underset{i\to\infty}{\longrightarrow} A(q). \tag{28}$$

An ∞-sequence of curves for which this limit is attained is then said *optimal*.

It is known [6] that $A(q) \leq q^{1/2} - 1$, with equality when q is a square [9][26].

Corollary 8. *Suppose $A(q) > 1$. Let r and r' be two positive real numbers such that*

$$r + r' \leq 1 - \frac{1}{A(q)}. \tag{29}$$

Then there exists a sequence of pairs of mutually intersecting codes (C, C') over $\mathbb{F}_q$, of length going to infinity, and of rates at least asymptotically (r, r').

Proof. Let X_i be curves forming an optimal sequence over $\mathbb{F}_q$. Let $G_i = \sum_{P \in X_i(\mathbb{F}_q)} P$ be the sum of all rational points in X_i. Since $A(q) > 1$, one has $n_i = \deg(G_i) = |X_i(\mathbb{F}_q)| > g_i$ for i big enough. Let m_i be a sequence of integers such that $\frac{m_i}{n_i} \longrightarrow r + \frac{1}{A(q)}$ as i goes to infinity (hence $g_i \leq m_i < n_i$ if i is big enough). Let D_i be an arbitrary divisor of degree m_i on X_i, and apply the preceding Proposition 7. This gives mutually intersecting codes (C_i, C'_i), where the rate of C_i is at least

$$\frac{m_i + 1 - g_i}{n_i} \underset{i\to\infty}{\longrightarrow} r + \frac{1}{A(q)} - \frac{1}{A(q)} = r \tag{30}$$

and the rate of C'_i is at least

$$\frac{n_i - m_i}{n_i} \underset{i\to\infty}{\longrightarrow} 1 - r - \frac{1}{A(q)} \geq r' \tag{31}$$

as claimed. □

Remark that for $r = r'$, this last corollary gives a family of mutually intersecting codes (C, C') of asymptotic rate $\frac{1}{2} - \frac{1}{2A(q)}$. This can be seen as a weak version of Theorem 2, which asserts that this can be done with $C = C'$ (but with more restrictive conditions on q).

3 The construction

From now on K is assumed to be a *perfect* field.

The main technical tool in the proof of Theorem 2 will be the following "higher version" of Lemma 6:

Lemma 9. *Let X be a curve over K of genus g, and let A be a divisor on X with $\deg(A) \leq g-3$ and*

$$l(A) = 0. \tag{32}$$

Then for all points $P \in X(K)$ except perhaps for at most $4g$ of them, we have

$$l(A+2P) = 0. \tag{33}$$

Proof. We can assume $|X(K)| > 4g \geq g$, otherwise there is nothing to prove. Then, thanks to Lemma 6, successively adding points to A, we can find a divisor $A' \geq A$ with $\deg(A') = g-3$ and $l(A') = 0$. Then for any $P \in X(K)$ with $l(A+2P) > 0$, we also have $l(A'+2P) > 0$. So we can replace A with A', that is, it suffices to prove Lemma 9 with $\deg(A) = g-3$. In turn, by Riemann-Roch, setting $B = \Omega - A$ where Ω is a canonical divisor on X, this is equivalent to the following statement:

If B is a divisor on X with $\deg(B) = g+1$ and $l(B) = 2$, then there are at most $4g$ points $P \in X(K)$ with $l(B-2P) > 0$.

Replacing B by a linearly equivalent divisor, we can suppose $B \geq 0$. Let then $\{1, f\}$ be a basis of $\mathcal{L}(B)$. We will conclude by a degree argument on the differential form df.

First we claim that df is non-zero. If $\operatorname{char} K = 0$ this is true because f is non-constant. If $\operatorname{char} K = p > 0$ then, since K is assumed perfect, $df = 0$ means $f = h^p$ for some $h \in K(X)$. But then $h \in \mathcal{L}(\frac{1}{p}B) \subset \mathcal{L}(B)$ and $\{1, h, f\}$ are linearly independent in $\mathcal{L}(B)$, contradicting our hypothesis $l(B) = 2$.

Let $\mathcal{S} = \{\, P \in X(K) \mid l(B-2P) > 0 \,\}$. We have to show $|\mathcal{S}| \leq 4g$.

Now if P is a closed point (of arbitrary degree) in X, we are in one of these four mutually exclusive situations:

(i) $P \notin \mathcal{S} \cup \operatorname{Supp}(B)$. Then $v_P(f) \geq 0$, and $v_P(df) \geq 0$.

(ii) $P \in \operatorname{Supp}(B) \setminus \mathcal{S}$. Then $v_P(B) \geq 1$, and $v_P(df) \geq v_P(f) - 1 \geq -v_P(B) - 1$ hence

$$v_P(df) \geq -2v_P(B). \tag{34}$$

(iii) $P \in \mathcal{S} \setminus \operatorname{Supp}(B)$. Consider the inclusions $\mathcal{L}(B-2P) \subset \mathcal{L}(B-P) \subset \mathcal{L}(B)$. By hypothesis $l(B) = 2$ and $l(B-2P) > 0$, and since $1 \in \mathcal{L}(B) \setminus \mathcal{L}(B-P)$, necessarily $\mathcal{L}(B-P) = \mathcal{L}(B-2P)$.

Now let $\alpha = f(P)$. Then $f - \alpha \in \mathcal{L}(B-P) = \mathcal{L}(B-2P)$, so $v_P(f-\alpha) \geq 2$, hence $v_P(d(f-\alpha)) \geq 1$. But since $df = d(f-\alpha)$ we conclude:

$$v_P(df) \geq 1. \tag{35}$$

(iv) $P \in \mathcal{S} \cap \mathrm{Supp}(B)$. By hypothesis $v_P(f) \geq -v_P(B)$ and $v_P(B) \geq 1$. We claim it is impossible to have simultaneously $v_P(f) = -v_P(B)$ and $v_P(B) = 1$.

For if it were the case, then $f \in \mathcal{L}(B) \setminus \mathcal{L}(B-P)$ and $1 \in \mathcal{L}(B-P) \setminus \mathcal{L}(B-2P)$, so all inclusions $\mathcal{L}(B-2P) \subset \mathcal{L}(B-P) \subset \mathcal{L}(B)$ would be strict, contradicting $l(B) = 2$ and $l(B-2P) > 0$.

So one (at least) of the inequalities $v_P(f) \geq -v_P(B)$ and $v_P(B) \geq 1$ is strict. Thus $v_P(df) \geq v_P(f) - 1 > -2v_P(B)$, that is:

$$v_P(df) \geq -2v_P(B) + 1. \tag{36}$$

Now summing these inequalities we find

$$2g - 2 = \deg(\mathrm{div}\, df) = \sum_P v_P(df) \deg(P) \geq -2\deg(B) + |\mathcal{S}| \tag{37}$$

and since $\deg(B) = g + 1$ this gives $|\mathcal{S}| \leq 4g$ as claimed. □

Proposition 10. *Let X be a curve over K of genus g, and suppose*

$$|X(K)| > 4g. \tag{38}$$

Let G be a divisor on X, of degree $n = \deg(G) \in \mathbb{Z}$. Then there exists a divisor D on X of degree $\deg(D) = \left\lfloor \frac{n+g-1}{2} \right\rfloor$ (or equivalently: $g-2 \leq \deg(2D-G) < g$), such that

$$l(2D - G) = 0. \tag{39}$$

Proof. For $0 \leq i \leq N = \left\lfloor \frac{n+g-1}{2} \right\rfloor - \left\lfloor \frac{n-1}{2} \right\rfloor$ we construct divisors D_i such that

$$\deg(D_i) = i + \left\lfloor \frac{n-1}{2} \right\rfloor \qquad \text{and} \qquad l(2D_i - G) = 0 \tag{40}$$

iteratively as follows:

- Start with any divisor D_0 of degree $\left\lfloor \frac{n-1}{2} \right\rfloor$, hence $\deg(2D_0 - G) < 0$ and $l(2D_0 - G) = 0$ as asked. For example take $P_0 \in X(K)$ and set $D_0 = \left\lfloor \frac{n-1}{2} \right\rfloor P_0$ — remark that $X(K)$ is non-empty, because of (38).
- Suppose up to some $i < N$, we have a divisor D_i satisfying (40). The divisor $A = 2D_i - G$ then satisfies $-2 \leq \deg(A) < g - 2$ and $l(A) = 0$, so by (38) and Lemma 9 we can find $P \in X(K)$ such that $l(A + 2P) = 0$. Then we set $D_{i+1} = D_i + P$, and D_{i+1} satisfies (40).
- This ends when $i = N$, and we can set $D = D_N$.

□

Remark that the construction given in the proof involves roughly $g/2$ iterations, and each step requires testing at most $4g+1$ points. So, as soon as a curve of genus g, as well as sufficiently many of its rational points, and the various Riemann-Roch spaces $\mathcal{L}(A)$, can be computed in time polynomial in g, then the overall construction will be polynomial in g.

Corollary 11. *Let X be a curve over K of genus g, such that $|X(K)| > 4g$. Let n be an integer such that $g < n \leq |X(K)|$. Then there exists a linear intersecting code C over K, of length n and dimension*

$$\dim C \;\geq\; \left\lfloor \frac{n+g-1}{2} \right\rfloor + 1 - g \;\geq\; \frac{n-g}{2}. \tag{41}$$

Proof. Let $G = P_1 + \cdots + P_n$ for pairwise distinct $P_i \in X(K)$. The proposition gives a divisor D on X of degree $\deg(D) = \lfloor \frac{n+g-1}{2} \rfloor < n$ with $l(2D-G) = 0$. The conclusion then follows from Theorem 4, with $C = C(G, D)$. □

We can now proceed with:

Proof of Theorem 2. Let X_i be curves forming an optimal sequence over $\mathbb{F}_q$, let g_i be the genus of X_i, and let G_i be the sum of all points in $X_i(\mathbb{F}_q)$, so $n_i = \deg(G_i) = |X_i(\mathbb{F}_q)|$. By definition we have $g_i \to \infty$ and $n_i/g_i \to A(q) > 4$, so $n_i > 4g_i$ if i is big enough. The preceding corollary then gives a linear intersecting code C_i over $\mathbb{F}_q$ of rate at least

$$\frac{1 - g_i/n_i}{2} \underset{i\to\infty}{\longrightarrow} \frac{1}{2} - \frac{1}{2A(q)} \tag{42}$$

as asked.

If $q \geq 25$ is a square, then $A(q) = q^{1/2} - 1$, and the conclusion follows, except perhaps for $q = 25$, $A(q) = 4$. But in this last particular case, we know that the sequence of modular curves $X_0(11\ell)$, for $\ell \geq 13$ prime, has genus $g_\ell = \ell$ and number of points $|X_0(11\ell)(\mathbb{F}_{25})| \geq 4\ell + 4 > 4g_\ell$, and we conclude in the same way. □

As regards constructiveness issues in this last proof, note that when q is a square, such optimal sequences are known explicitly (see for example [8]), and all computations in the proposition can be made in polynomial time, hence the overall construction can be made in polynomial time (although perhaps with constants and exponents too big to be really useful in practice).

Remark 12. We finish by noting two possible improvements on Theorem 2.

(i) In fact the hypothesis $A(q) > 4$ (or $q = 25$, $A(q) = 4$) in Theorem 2 is not optimal. This constant 4 comes from Lemma 9, and it turns out that the estimation in this lemma (as well as the one in Lemma 6, by the way) can be slightly improved, as done in [19] (the proof is more technically involved).

From this stronger version of the lemma one can show that the conclusion in Theorem 2 holds already when $A(q) \geq 4 - \frac{12q^2-4}{q^4+2q^2-1}$ (see [19] for more details).

Clearly this improvement is small, not to say unimpressive, and for the application to Theorem 1, we only need the case $q = 121$, $A(q) = 10$, so we can leave such refinements apart. Nevertheless, further relaxing of the condition on q in Theorem 2 could have interest by itself.

(ii) Theorem 2 is concerned only in improving the case $s = 2$ of Xing's bound [27] (we will keep his notations, so our R_q becomes $R_q(2)$), since this is all we need for Theorem 1 again. However, following [17], it is natural to conjecture that, for any s, and maybe under suitable conditions on q,

$$R_q(s) \geq \frac{1}{s} - \frac{1}{A(q)} + \frac{1}{sA(q)}. \tag{43}$$

If one tries to prove (43) with a method similar to the one given here, one will construct inductively some divisors A of controlled degree and dimension $l(A) = 0$, and the main point will be to show that, given sufficiently many points, there is one of them, say P, such that $l(A+sP) = 0$ (of which Lemma 6 is the case $s = 1$ and Lemma 9 the case $s = 2$). Equivalently (see e.g. [13] or [25]) one has to prove that $B = \Omega - A$ has order sequence at P starting with $\epsilon_0(P) = 0, \ldots, \epsilon_{s-1}(P) = s - 1$. A necessary condition for this to be possible, is that B has *generic* order sequence starting with $\epsilon_0 = 0, \ldots, \epsilon_{s-1} = s - 1$. If this holds, the existence of P can be derived from a Plücker formula ([13], Theorem 9).

For $s = 2$, it is known that any complete linear system has generic order sequence starting with $\epsilon_0 = 0$ and $\epsilon_1 = 1$. In our situation, this is equivalent to the non-vanishing of df established during the proof of Lemma 9 — and then the proof of Lemma 9 proceeds with a variant of the Plücker formula suitable for our particular case (classically, this relies on Wronskians; in the proof given here, the Wronskian is just df).

Unfortunately, for $s \geq 3$, not all divisors B have generic order sequence starting with $\epsilon_0 = 0, \ldots, \epsilon_{s-1} = s - 1$. While this is known to hold for "most" divisors [16], it might be difficult to ensure that it is so for the particular divisors constructed in an inductive procedure such as ours.

References

[1] N. Alon and J. H. Spencer, *The probabilistic method, 2nd ed.*, Wiley-Interscience, 2000.

[2] S. R. Blackburn, *Frameproof codes*, SIAM J. Discrete Math. **16** (2003) 499–510.

[3] D. Boneh and J. Shaw, *Collusion-secure fingerprinting for digital data*, IEEE Trans. Inform. Theory **44** (1998) 1897–1905.

[4] G. Cohen and A. Lempel, *Linear intersecting codes*, Discr. Math. **56** (1985) 35–43.

[5] G. Cohen and H. G. Schaathun, *Asymptotic overview on separating codes*, Reports in Informatics **248**, Univ. Bergen, 2003.

[6] V. G. Drinfeld and S. G. Vladut, *Number of points of an algebraic curve*, Funct. Anal. **17** (1983) 53–54.

[7] A. D. Friedman, R. L. Graham, and J. D. Ullman, *Universal single transition time asynchronous state assignments*, IEEE Trans. Comput. **18** (1969) 541–547.

[8] A. Garcia and H. Stichtenoth, *A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound*, Invent. Math. **121** (1995) 211–222.

[9] Y. Ihara, *Some remarks on the number of rational points of algebraic curves over finite fields*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **28** (1981) 721–724.

[10] J. Komlós, unpublished, cited in [15].

[11] J. Körner, *On the extremal combinatorics of the Hamming space*, J. Combin. Theory Ser. A **71** (1995) 112–126.

[12] A. Krasnopeev and Yu. L. Sagalovich, "The Kerdock codes and separating systems", in: *ACCT-8, Tsarskoe Selo (St Petersbourg) 2002*, 165–167.

[13] D. Laksov, *Wronskians and Plücker formulas for linear systems on curves*, Ann. Sci. École Norm. Sup. **17** (1984) 45–56.

[14] A. Lempel and S. Winograd, *A new approach to error-correcting codes*, IEEE Trans. Inform. Theory **23** (1977) 503–508.

[15] D. Miklós, *Linear binary codes with intersection properties*, Discr. Appl. Math. **9** (1984) 187–196.

[16] A. Neeman, *Weierstrass points in characteristic p*, Invent. Math. **75** (1984) 359–376.

[17] H. Randriam, "Hecke operators with odd determinant and binary frame-proof codes beyond the probabilistic bound?", in *Proceedings of 2010 IEEE Information Theory Workshop (ITW 2010 Dublin)*, to appear.

[18] H. Randriambololona, preliminary version of the present article: $(2,1)$-*separating systems beyond the probabilistic bound*, April 2011, available online at: `http://arxiv.org/abs/1010.5764v6`

[19] H. Randriambololona, *Diviseurs de la forme $2D - G$ sans sections et rang de la multiplication dans les corps finis*, preprint, available online at: `http://arxiv.org/abs/1103.4335`

[20] A. Rényi, *On random generating elements of a finite Boolean algebra*, Acta Sci. Math. Szeged **22** (1961) 75–81.

[21] Yu. L. Sagalovich, *Cascade codes of automata states*, Probl. Peredachi Inf. **14** (1978) 77–85.

[22] Yu. L. Sagalovich and A. G. Chilingarjan, *Separating systems and new scopes of its application*, Information Processes **9** (2009) 225–248.

[23] H. Stichtenoth, *Algebraic function fields and codes*, Universitext, Springer-Verlag, 1993.

[24] D. R. Stinson and R. Wei, *Combinatorial properties and constructions of traceability schemes and frameproof codes*, SIAM J. Discr. Math. **11** (1998) 41–53.

[25] K.-O. Stöhr and F. Voloch, *Weierstrass points and curves over finite fields*, Proc. London Math. Soc. **52** (1986) 1–19.

[26] M. A. Tsfasman, S. G. Vladut, and T. Zink, *Modular curves, Shimura curves, and Goppa codes better than Varshamov-Gilbert bound*, Math. Nachr. **109** (1982) 21–28.

[27] C. Xing, *Asymptotic bounds on frameproof codes*, IEEE Trans. Inform. Theory **48** (2002) 2991–2995.