On the minimum distance of elliptic curve codes

Jiyou Li
Department of Mathematics
Shanghai Jiao Tong University
Shanghai, P.R.China
Email: lijiyou@sjtu.edu.cn

Daqing Wan
Department of Mathematics
University of California, Irvine
CA 92697-3875, USA
Email: dwan@math.uci.edu

Jun Zhang
School of Mathematical Sciences
Capital Normal University
Beijing 100048, P.R.China
Email: junz@cnu.edu.cn

***Abstract*—Computing the minimum distance of a linear code is one of the fundamental problems in algorithmic coding theory. Vardy [1] showed that it is an NP-hard problem for general linear codes. In practice, one often uses codes with additional mathematical structure, such as cyclic codes and algebraic geometry (AG) codes, etc. In this paper, we study the minimum distance of a family of AG codes. For AG codes of genus 0 (generalized Reed-Solomon codes), the minimum distance has a simple explicit formula. An interesting result of Cheng [2] says that the minimum distance problem is already NP-hard (under RP-reduction) for general elliptic curve codes (ECAG codes, or AG codes of genus 1). In this paper, we show that the minimum distance of ECAG codes also has a simple explicit formula if the evaluation set is suitably large (at least $2/3$ of the group order). Our method is purely combinatorial and based on a new sieving technique from Li-Wan [3].**

I. Introduction

Let $\mathbb{F}_q^n$ be the n-dimensional vector space over the finite field $\mathbb{F}_q$ with q elements. For any vector $x = (x_1, x_2, \cdots, x_n) \in \mathbb{F}_q^n$, the *Hamming weight* $\mathrm{Wt}(x)$ of x is defined to be the number of non-zero coordinates, i.e.,

$$\mathrm{Wt}(x) = |\{i \,|\, 1 \leqslant i \leqslant n,\, x_i \neq 0\}| \ .$$

A *linear* $[n,k]$ *code* C is a k-dimensional linear subspace of $\mathbb{F}_q^n$. The *minimum distance* $d(C)$ of C is the minimum Hamming weight of all non-zero vectors in C, i.e.,

$$d(C) = \min\{\mathrm{Wt}(c) \,|\, c \in C \setminus \{0\}\} \ .$$

A linear $[n,k]$ code $C \subseteq \mathbb{F}_q^n$ is called a $[n,k,d]$ linear code if C has minimum distance d. A well-known trade-off between the parameters of a linear $[n,k,d]$ code is the Singleton bound which states that

$$d \leqslant n - k + 1 \ .$$

An $[n,k,d]$ code is called a *maximum distance separable* (MDS) code if the equality above holds, i.e., $d = n - k + 1$.

The minimum distance of a linear code determines the ability of detecting and correcting of the code. Computing the minimum distance of a linear code is one of the most important problems in algorithmic coding theory. It was proved to be **NP**-hard for general linear codes in [1]. The gap version of the problem was also shown to be **NP**-hard in [4]. And the same paper showed that approximating the minimum distance of a linear code cannot be achieved in randomized polynomial time to the factor $2^{\log^{1-\epsilon} n}$ unless $\mathrm{NP} \subseteq \mathrm{RTIME}(2^{\mathrm{polylog}(n)})$. In [5], Cheng and the second author derandomized the reduction and showed there is no deterministic polynomial time algorithm to approximate the minimum distance to any constant factor unless $\mathrm{NP} = \mathrm{P}$. And they proved that approximating the minimum distance of a linear code cannot be achieved in deterministic polynomial time to the factor $2^{\log^{1-\epsilon} n}$ unless $\mathrm{NP} \subseteq \mathrm{RTIME}(2^{\mathrm{polylog}(n)})$.

Despite the above complexity results, it is more interesting to compute the minimum distance of linear codes that are used in practical applications. An important class of such codes is algebraic geometry (AG) codes with parameters $[n,k,d]$ as defined in Section 4. The minimum distance of such AG codes from algebraic curves of genus g is known to satisfy the inequality

$$n - k - g + 1 \leq d \leq n - k + 1.$$

In the simplest case $g = 0$, i.e., generalized Reed-Solomon codes, the minimum distance has the simple formula $d = n - k + 1$. In the next simplest case $g = 1$, either $d = n - k$ or $d = n - k + 1$, and Cheng [2] showed that determining the minimum distance of ECAG codes between the two options is **NP**-hard under **RP**-reduction. For genus $g \geq 2$, there is no such complexity result so far. But it is believed to be an **NP**-hard problem as well.

We are interested in positive results for determining the minimum distance of ECAG codes. It was shown in [2], and also in [6] from a different aspect, that computing the minimum distance of an ECAG code is equivalent to a subset sum problem (SSP) in the group of rational points on the elliptic curve. We now make this more precise.

Let E be an elliptic curve over the finite field $\mathbb{F}_q$. Let G be the group of $\mathbb{F}_q$-rational points on the elliptic curve E. The Hasse bound shows that $||G|-(q+1)| \leq 2\sqrt{q}$. Let $D \subseteq G$ be a nonempty subset of cardinality n, which will be our evaluation set for ECAG code. For a positive integer $1 \leq k \leq n < |G|$ and element $b \in G$, let $N(k,b,D)$ be the number of k-subsets $T \subseteq D$ such that $\sum_{x \in T} x = b$. The counting version of the k-subset sum problem for the pair (G,D) is to compute $N(k,b,D)$. The minimum distance of the ECAG $[n,k]$-code is equal to $n-k$ if and only if the number $N(k,b,D)$ is positive. This k-subset sum problem is in general **NP**-hard if the evaluation set D is small. On the other hand, the dynamic programming method implies that there is a polynomial time algorithm to compute $N(k,b,D)$ if $n = |D|$ is large, say, $n = |G|^\delta$ for some constant $\delta > 0$.

In this paper, we obtain an asymptotic formula for $N(k,b,D)$ if $n = |D|$ is suitably large, say, $|D| > (\frac{2}{3} + \epsilon)|G|$. As an application, we show that if the cardinality n of the evaluation set is suitably large (at least $2/3$ of the group size),

then the minimum distance of an ECAG code $[n,k]$ is always $n-k$. We conjecture that the condition $|D| > (\frac{2}{3}+\epsilon)|G|$ in our results can be improved to $|D| > (\frac{1}{2}+\epsilon)|G|$. Our main technical tool is the sieve method of Li-Wan [7].

Let $\widehat{G}$ be the group of additive characters of G. Note that $\widehat{G}$ is isomorphic to G. Denote the amplitude by

$$\Phi(D) = \max_{\chi\in\widehat{G},\chi \text{ is non-trivial}} \left|\sum_{a\in D}\chi(a)\right|.$$

Our main technical result is the following asymptotic formula for $N(k,b,D)$.

Theorem 1.1: Notations as above. We have

$$\begin{aligned}\left|N(k,b,D) - |G|^{-1}\tbinom{n}{k}\right| \;&\le \tfrac{|S|}{|G|}\tbinom{\Phi(D)+k-1}{k} + \tfrac{1}{|G|}\tbinom{\frac{n+\Phi(D)}{2}}{k} \\ &+\tfrac{1}{|G|}\textstyle\sum_{\substack{2<d\le k\\ d\mid \exp(G)}} \phi(d)\tbinom{\frac{n+\Phi(D)}{d}+k-1}{k},\end{aligned}$$

where S is the set of characters in $\widehat{G}$ which have order greater than k and $\exp(G)$ is the exponent of G.

We apply this theorem to determine the minimum distance of ECAG codes (for details see Section IV) and obtain

Theorem 1.2: Suppose that $n \ge (\frac{2}{3}+\epsilon)q$ and $q > \frac{4}{\epsilon^2}$, where ϵ is positive. There is a positive constant C_ϵ such that if $C_\epsilon \ln q < k < n - C_\epsilon \ln q$, then ECAG codes $[n,k]$ have the deterministic minimum distance $n-k$.

In other words, if an elliptic curve code $[n,k]$ over $\mathbb{F}_q$ is an MDS code, then the length n should not exceed $(\frac{2}{3}+\epsilon)q$. Munuera [8] got an upper bound $\frac{q+1}{2}+\sqrt{q}+k$ for the length of MDS elliptic curve codes. For fixed k, when q is sufficiently large, Munuera's bound tends to about $\frac{1}{2}q$. But for large k, saying cq, Munuera's bound becomes looser and looser when c becomes larger and larger, and it turns to the MDS conjecture when c is close to $\frac{1}{2}$. However, compared with Munuera's bound, our result holds for almost all k in the range $[1,n]$. Especially, our bound performances better than Munuera's bound when k is large.

If we allow the length of the codes to be larger, we then have a better bound on k.

Theorem 1.3: If $n \ge q+2$, then for $q > 64$ and $3 < k < q-1$, then ECAG $[n,k]$ codes have the deterministic minimum distance $n-k$.

Note that one can check the cases $q \le 64$ by a computer search, we have a complete result for the minimum distance of the ECAG code $[n,k]$ if $n \ge q+2$. This gives a new proof of MDS conjecture on ECAG codes, in a purely combinatoric method. For recent progress on MDS conjecture on general codes and related problems, please see [9].

This paper is organized as follows. Section 2 recalls the sieve method of Li-Wan. Section 3 uses the sieve method to get an estimate of counting subset sum problems on any large subset of the rational point group of an elliptic curve. And Section 4 describes the relation between minimum distance of ECAG codes and subset sum problems on the evaluation set of the ECAG code. The main result of this paper then follows.

II. A DISTINCT COORDINATE SIEVING FORMULA

For the purpose of the proof, we introduce a sieving formula discovered by Li-Wan [3], which significantly improves the classical inclusion-exclusion sieve in many interesting cases. We cite it here without any proof. For details and related applications, we refer to [3], [7].

Before we present the sieving formula, we introduce some notations valid for the whole paper. Let D be an alphabet set, X a finite set of vectors of length k over D. Denote $\overline{X} = \{(x_1,x_2,\cdots,x_k) \in X \mid x_i \ne x_j, \forall i \ne j\}$ the pairwise distinct component subset. Let S_k be the symmetric group on $\{1,2,\cdots,k\}$. For $\tau \in S_k$, the $sign$ function is defined to be $sign(\tau) = (-1)^{k-l(\tau)}$, where $l(\tau)$ is the number of cycles of τ including the trivial cycles which have length 1. Let $\tau = (i_1 i_2 \cdots i_{a_1})(j_1 j_2 \cdots j_{a_2})\cdots(l_1 l_2 \cdots l_{a_s})$ with $1 \le a_i, 1 \le i \le s$ be any permutation, denote the τ-symmetric subset

$$\begin{aligned}X_\tau = \;&\big\{(x_1,\ldots,x_k) \in X \,|\, x_{i_1} = \cdots = x_{i_{a_1}}, \cdots, \\ &x_{l_1} = \cdots = x_{l_{a_s}}\big\}.\end{aligned} \tag{1}$$

Let $f(x_1,x_2,\ldots,x_k)$ be a complex valued function defined on X. Denote the distinct sum

$$F = \sum_{x\in\overline{X}} f(x_1,x_2,\ldots,x_k),$$

and the τ-symmetric sum

$$F_\tau = \sum_{x\in X_\tau} f(x_1,x_2,\ldots,x_k).$$

We now state our sieve formula. We remark that there are many other interesting corollaries of this formula. For interested reader we refer to [3].

Theorem 2.1: Let F and F_τ be defined as above. Then

$$F = \sum_{\tau\in S_k} sign(\tau) F_\tau. \tag{2}$$

For $\tau \in S_k$, denote by $\overline{\tau}$ the conjugacy class determined by τ whose elements are permutations conjugate to τ. Conversely, for given conjugacy class $\overline{\tau} \in C_k$, denote by τ a representative permutation of this class. For convenience we usually identify these two symbols. Since two permutations in S_k are conjugate if and only if they have the same type of cycle structure (up to the order), C_k is exactly the set of all partitions of k.

The symmetric group S_k acts on D^k naturally by permuting coordinates. That is, for $\tau \in S_k$ and $x = (x_1,x_2,\ldots,x_k) \in D^k$, $\tau\circ x = (x_{\tau(1)}, x_{\tau(2)}, \ldots, x_{\tau(k)})$. A subset X in D^k is said to be *symmetric* if for any $x \in X$ and any $\tau \in S_k$, $\tau\circ x \in X$. In particular, if X is symmetric and f is a symmetric function under the action of S_k, we then have the following simpler formula for (2).

Proposition 2.2: Let C_k be the set of conjugacy classes of S_k. If X is symmetric and f is symmetric, then

$$F = \sum_{\tau\in C_k} sign(\tau) C(\tau) F_\tau, \tag{3}$$

where $C(\tau)$ is the number of permutations conjugate to τ.

For the purpose of evaluating the above summation, we need several combinatorial formulas. A permutation $\tau \in S_k$ is said to be of type $(c_1, c_2, \cdots, c_k)$ if τ has exactly c_i cycles of length i. Note that $\sum_{i=1}^{k} ic_i = k$. Let $N(c_1, c_2, \ldots, c_k)$ be the number of permutations in S_k of type $(c_1, c_2, \ldots, c_k)$ and it is well-known that

$$N(c_1, c_2, \ldots, c_k) = \frac{k!}{1^{c_1} c_1! 2^{c_2} c_2! \cdots k^{c_k} c_k!}.$$

Lemma 2.3: Define the generating function

$$C_k(t_1, t_2, \ldots, t_k) = \sum_{\sum ic_i = k} N(c_1, c_2, \ldots, c_k) t_1^{c_1} t_2^{c_2} \cdots t_k^{c_k}.$$

If $t_1 = t_2 = \cdots = t_k = q$, then we have

$$\begin{aligned} C_k(q, q, \ldots, q) &= \textstyle\sum_{\sum ic_i = k} N(c_1, c_2, \ldots, c_k) q^{c_1} q^{c_2} \cdots q^{c_k} \\ &= (q + k - 1)_k \end{aligned}$$

In another case, if $t_i = q$ for $d \mid i$ and $t_i = s$ for $d \nmid i$, then we have

$$\begin{aligned} & C_k(\overbrace{s, \cdots, s}^{d-1}, q, \overbrace{s, \cdots, s}^{d-1}, q, \cdots) \\ = & \textstyle\sum_{\sum ic_i = k} N(c_1, c_2, \cdots, c_k) q^{c_1} q^{c_2} \cdots s^{c_d} q^{c_{d+1}} \cdots \\ = & k! \textstyle\sum_{i=0}^{\lfloor k/d \rfloor} \binom{\frac{q-s}{d} + i - 1}{\frac{q-s}{d} - 1} \binom{s + k - di - 1}{s - 1} \\ \leq & k! \binom{s + k + (q-s)/d - 1}{k}. \end{aligned}$$

III. Subset Sum Problem in a Subset of the Rational Point Group

Lemma 3.1 (Hasse-Weil Bound): Let E be an elliptic curve over the finite field $\mathbb{F}_q$. Then the number of rational points on E has the following estimate

$$|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}.$$

It is well-known that the rational points $E(\mathbb{F}_q)$ form a finite group and more precisely, it has the structure $G = E(\mathbb{F}_q) \cong \mathbb{Z}/n_1 \times \mathbb{Z}/n_2$ for some $n_1 | n_2$. By Lemma 3.1, G has order $q + 1 + c\sqrt{q}$, with $|c| \leq 2$. Denote by $\exp(G)$ the exponent of G. Let $D \subseteq G$ be a nonempty subset of cardinality n. Let $\widehat{G}$ be the group of additive characters of G with the trivial character χ_0. Note that $\widehat{G}$ is isomorphic to G. Denote the partial character sum $s_\chi(D) = \sum_{a \in D} \chi(a)$ and the amplitude $\Phi(D) = \max_{\chi \in \widehat{G}, \chi \neq \chi_0} |s_\chi(D)|$. Let $N(k, b, D)$ be the number of k-subsets $T \subseteq D$ such that $\sum_{x \in S} x = b$. In the following theorem we will give an asymptotic bound for $N(k, b, D)$ which ensures $N(k, b, D) > 0$ when $G - D$ is not too large compared with G.

Theorem 3.2: Let $N(k, b, D)$ be defined as above.

$$\begin{aligned} \left| N(k, b, D) - |G|^{-1} \binom{n}{k} \right| \leq & \frac{|S|}{|G|} \binom{\Phi(D) + k - 1}{k} + \frac{1}{|G|} \binom{\frac{n + \Phi(D)}{2}}{k} \\ & + \frac{1}{|G|} \sum_{\substack{2 < d \leq k \\ d \mid \exp(G)}} \phi(d) \binom{\frac{n + \Phi(D)}{d} + k - 1}{k}, \end{aligned} \tag{4}$$

where S is the set of characters which has order greater than k.

Proof: Let $X = D \times D \times \cdots \times D$ be the Cartesian product of k copies of D. Let $\overline{X} = \left\{ (x_1, x_2, \ldots, x_k) \in D^k \mid x_i \neq x_j, \ \forall i \neq j \right\}$. It is clear that $|X| = n^k$ and $|\overline{X}| = (n)_k$. We have

$$\begin{aligned} & k! N(k, b, D) \\ = & \frac{1}{|G|} \sum_{(x_1, x_2, \ldots x_k) \in \overline{X}} \sum_{\chi \in \widehat{G}} \chi(x_1 + x_2 + \cdots + x_k - b) \\ = & \frac{1}{|G|} (n)_k + \frac{1}{|G|} \sum_{\chi \neq \chi_0} \sum_{(x_1, x_2, \cdots x_k) \in \overline{X}} \chi(x_1) \cdots \chi(x_k) \chi^{-1}(b) \\ = & \frac{1}{|G|} (n)_k + \frac{1}{|G|} \sum_{\chi \neq \chi_0} \chi^{-1}(b) \sum_{(x_1, x_2, \ldots x_k) \in \overline{X}} \textstyle\prod_{i=1}^{k} \chi(x_i). \end{aligned}$$

Denote $f_\chi(x) = f_\chi(x_1, x_2, \ldots, x_k) = \prod_{i=1}^{k} \chi(x_i)$. For $\tau \in S_k$, let

$$F_\tau(\chi) = \sum_{x \in X_\tau} f_\chi(x) = \sum_{x \in X_\tau} \prod_{i=1}^{k} \chi(x_i),$$

where X_τ is the τ-symmetric subset which is defined as in (1). Obviously X is symmetric and $f_\chi(x_1, x_2, \ldots, x_k)$ is normal on X. Applying (3) in Corollary 2.2, we get

$$\begin{aligned} & k! N(k, b, D) \\ = & \frac{1}{|G|} (n)_k + \frac{1}{|G|} \sum_{\chi \neq \chi_0} \chi^{-1}(b) \sum_{\tau \in C_k} sign(\tau) C(\tau) F_\tau(\chi), \end{aligned}$$

where C_k is the set of conjugacy classes of S_k, $C(\tau)$ is the number of permutations conjugate to τ. If τ is of type $(c_1, c_2, \ldots, c_k)$, then

$$\begin{aligned} F_\tau(\chi) &= \sum_{x \in X_\tau} \prod_{i=1}^{k} \chi(x_i) \\ &= \sum_{x \in X_\tau} \prod_{i=1}^{c_1} \chi(x_i) \prod_{i=1}^{c_2} \chi^2(x_{c_1 + 2i}) \\ &\quad \cdots \prod_{i=1}^{c_k} \chi^k(x_{c_1 + c_2 + \cdots + ki}) \\ &= \prod_{i=1}^{k} (\sum_{a \in D} \chi^i(a))^{c_i} \\ &= n^{\sum c_i m_i(\chi)} s_\chi(D)^{\sum c_i (1 - m_i(\chi))}, \end{aligned}$$

where $m_i(\chi) = 1$ if $\chi^i = 1$ and otherwise $m_i(\chi) = 0$.

Now suppose $ord(\chi) = d$ with $d \mid n_1 n_2$. Note that $C(\tau) = N(c_1, c_2, \ldots, c_k)$. In the case $d = 2$, $s_\chi(D)$ is an integer. Applying Lemma 2.3, we have

$$\begin{aligned} & \sum_{\tau \in C_k} sign(\tau) C(\tau) F_\tau(\chi) \\ = & (-1)^k \sum_{\tau \in C_k} C(\tau) (-n)^{\sum c_i m_i(\chi)} (-s_\chi(D))^{\sum c_i (1 - m_i(\chi))} \\ = & (-1)^k k! \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{\frac{-n + s_\chi(D)}{2} + i - 1}{i} \binom{-s_\chi(D) + k - 2i - 1}{k - 2i} \\ = & k! \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{\frac{n - s_\chi(D)}{2}}{i} \binom{s_\chi(D)}{k - 2i} \\ \leq & k! \binom{\frac{n + \Phi(D)}{2}}{k}. \end{aligned}$$

The last inequality in the case $s_\chi(D) > 0$ is from the identity

$$\sum_{i=0}^{k} \binom{a}{i} \binom{b}{k - i} = \binom{a + b}{k}.$$

In the case $s_\chi(D) < 0$, since the summation has alternative signs, the inequality follows from a simple combinatorial argument.

In the case $3 \le d \le k$, since $|s_\chi(D)| \le \Phi(D)$, we have

$$\begin{aligned}&\sum_{\tau\in C_k} sign(\tau)C(\tau)F_\tau(\chi)\\ \le&\sum_{\tau\in C_k} C(\tau)n^{\sum c_i m_i(\chi)}\Phi(D)^{\sum c_i(1-m_i(\chi))}\\ \le& k!\binom{\frac{n+\Phi(D)}{d}+k-1}{k}.\end{aligned}$$

Similarly, if $ord(\chi)$ is greater than k, then

$$\sum_{\tau\in C_k} sign(\tau)C(\tau)F_\tau(\chi) \le k!\binom{\Phi(D)+k-1}{k}.$$

Let S be the set of characters which have order greater than k. Summing over all nontrivial characters, we obtain

$$\begin{aligned}\left|N(k,b,D)-|G|^{-1}\tbinom{n}{k}\right| \le& \tfrac{|S|}{|G|}\tbinom{\Phi(D)+k-1}{k}+\tfrac{1}{|G|}\tbinom{\frac{n+\Phi(D)}{2}}{k}\\ &+\tfrac{1}{|G|}\sum_{\substack{2<d\le k\\ d|\exp(G)}}\phi(d)\tbinom{\frac{n+\Phi(D)}{d}+k-1}{k},\end{aligned}$$

where $\phi(d)$ is the number of characters in $\widehat{G}$ of order d. This completes the proof. ∎

Corollary 3.3: We have

$$\left|N(k,b,D)-|G|^{-1}\binom{n}{k}\right| \le \binom{M}{k},$$

where M is defined as

$$M=\max\left\{\binom{\Phi(D)+k-1}{k},\binom{\frac{n+\Phi(D)}{2}}{k},\binom{\frac{n+\Phi(D)}{d}+k-1}{k}\right\},$$

and d is the smallest nontrivial divisor of $|G|$ that is not equal to 2.

Corollary 3.4: Let $q \ge 64$ and $n = q+2$. For $6 \le k < q-1$, we have $N(k,b,D) > 0$ for every $b \in G$.

Proof: By symmetry it is sufficient to consider the case $3 \le k \le n/2$. To ensure $N(k,b,D) > 0$, by (4) it suffices to have

$$\begin{aligned}\binom{n}{k} >&|S|\binom{\Phi(D)+k-1}{k}+\binom{\frac{n+\Phi(D)}{2}}{k}\\ &+\sum_{\substack{2<d\le k\\ d|\exp(G)}}\phi(d)\binom{\frac{n+\Phi(D)}{d}+k-1}{k}.\end{aligned}$$

For a nontrivial character χ, $\sum_{g\in G}\chi(g)=0$ and it follows that $\Phi(D)=\Phi(G-D)<|G|-|D|\le 2\sqrt{q}+1$.

Since G is the product of at most two cyclic groups, by the definition of $\phi(d)$ we have $\phi(d)\le d^2-1$. For simplicity, set $K=k^3-2k^2-k+2$. For the case $k \le q^{1/3}$, it is sufficient to have

$$\begin{aligned}&\binom{q+2}{k}-(q+2\sqrt{q}-K)\binom{2\sqrt{q}+k-1}{k}\\ &-\binom{\frac{q+2+2\sqrt{q}}{2}}{k}-K\binom{\frac{q+2\sqrt{q}}{3}+k-1}{k}>0\end{aligned}$$

When $k=3$, one has

$$\begin{aligned}&125/216q^3-379/36q^{5/2}-589/18q^2\\ &+593/27q^{3/2}+149/2q+67/3q^{1/2}>0.\end{aligned}$$

Similarly, when $k=6$, one has $q\ge 64$. This is done by first taking $K=k^3-2k^2-k+2=140$, we solve that $q\ge 97$. But notice that now K should be ≤ 117. Then taking $K=117$, we solve $q\ge 79$. Iteratively, we can get $q\ge 64$ finally.

One checks that when $k \le q^{1/3}$ this function is unimodal on k. For $q^{1/3}<k<(q+2\sqrt{q})/6$, it then suffices to have

$$\binom{q+2}{k} > (q+2+2\sqrt{q})\binom{\frac{q+2+2\sqrt{q}}{2}}{k},$$

and for $(q+2\sqrt{q})/6 \le k \le (q+2)/2$,

$$\binom{q+2}{k} > (q+2+2\sqrt{q})\binom{\frac{q+2\sqrt{q}}{3}+k-1}{k}.$$

It follows from a simple asymptotic analysis and the proof is complete. ∎

A similar argument gives

Corollary 3.5: Suppose that $n \ge (\frac{2}{3}+\epsilon)q$ and $q > \frac{4}{\epsilon^2}$, where ϵ is positive. Then there is a positive constant C_ϵ such that $N(k,b,D)>0$ for every $b\in G$ provided $C_\epsilon \ln q < k < n - C_\epsilon \ln q$.

Proof: Similarly as above, we only consider the case $k \le n/2$. To ensure $N(k,b,D)>0$, by (4) it suffices to have

$$\begin{aligned}\binom{n}{k} >&|S|\binom{\Phi(D)+k-1}{k}+\binom{\frac{n+\Phi(D)}{2}}{k}\\ &+\sum_{\substack{2<d\le k\\ d|\exp(G)}}\phi(d)\binom{\frac{n+\Phi(D)}{d}+k-1}{k}.\end{aligned}$$

For a nontrivial character χ, $\sum_{g\in G}\chi(g)=0$ and it follows that $\Phi(D)=\Phi(G-D)<|G|-|D|\le(\frac{1}{3}-\epsilon)q+2\sqrt{q}+1$.

For small $k \le q/6$ it suffices to have

$$\binom{\frac{2q}{3}}{k}-(q+2\sqrt{q})\binom{\frac{q+2+2\sqrt{q}}{2}}{k}>0,$$

which holds when $k > C\ln q$ for some constant C.

For $q/6 < k \le n/2 = (\frac{1}{3}+\frac{\epsilon}{2})q$, it suffices to have

$$\binom{(\frac{2}{3}+\epsilon)q}{k} > (q+2\sqrt{q})\binom{(\frac{2}{3}+\frac{\epsilon}{2})q+\sqrt{q}}{k},$$

which holds when $q>\frac{4}{\epsilon^2}$ and $k > C_\epsilon \ln q$ for some constant C_ϵ. So the proof is complete. ∎

From the proof of the above corollary, if follows that

Corollary 3.6: Suppose $n \ge (\frac{2}{3}+\epsilon)q$, where ϵ is positive and $\epsilon \le 1/3$. When q is large enough (in application we need to use long length codes, so it is reasonable to assume q is large), then there is a positive constant C (independent of ϵ and q) such that $N(k,b,D)>0$ for every $b\in G$ provided $C\ln q < k < n - C\ln q$.

IV. Minimum Distance of Elliptic Codes and SSP

In this section, we discuss the relationship between the minimum distance of ECAG code and SSP on the group of rational points of the elliptic curve. Using the results in the previous section, our main theorems in Introduction follow automatically.

Let $E/\mathbb{F}_q$ be an elliptic curve over the finite field $\mathbb{F}_q$ with function field $\mathbb{F}_q(E)$. Let $E(\mathbb{F}_q)$ be the set of all $\mathbb{F}_q$-rational points on E. Suppose $D = \{P_1, P_2, \cdots, P_n\}$ is a proper subset of rational points $E(\mathbb{F}_q)$, and G is a divisor of degree k ($2g-2 < k < n$) with $\mathrm{Supp}(G) \cap D = \emptyset$. Without any confusion, we also write $D = P_1 + P_2 + \cdots + P_n$. Denote by $\mathscr{L}(G)$ the $\mathbb{F}_q$-vector space of all rational functions $f \in \mathbb{F}_q(E)$ with the principal divisor $\mathrm{div}(f) \geqslant -G$, together with the zero function (cf. [10]).

The functional AG code $C_{\mathscr{L}}(D, G)$ is defined to be the image of the following evaluation map:

$$ev : \mathscr{L}(G) \to \mathbb{F}_q^n;\ f \mapsto (f(P_1), f(P_2), \cdots, f(P_n)).$$

It is well-known that $C_{\mathscr{L}}(D, G)$ has parameters $[n, k, d]$ where the minimum distance d has two choices:

$$d = n - k, \ \text{ or } \ d = n - k + 1.$$

Suppose O is one of the $\mathbb{F}_q$-rational points on E . The set of rational points $E(\mathbb{F}_q)$ forms an abelian group with zero element O (for the definition for the sum of any two points, we refer to [11]), and it is isomorphic to the Picard group $\mathrm{div}^o(E)/\mathrm{Prin}(\mathbb{F}_q(E))$ where $\mathrm{Prin}(\mathbb{F}_q(E))$ is the subgroup consisting of all principal divisors. Denote by $\oplus$ and $\ominus$ the additive and minus operator in the group $E(\mathbb{F}_q)$, respectively.

Proposition 4.1 ([2], [6]): Let E be an elliptic curve over $\mathbb{F}_q$, $D = \{P_1, P_2, \cdots, P_n\}$ a subset of $E(\mathbb{F}_q)$ such that rational points (not necessarily distinct) $O, P \notin D$ and let $G = (k-1)O + P$ ($0 < k < n$). Endow $E(\mathbb{F}_q)$ a group structure with the zero element O. Then the AG code $C_{\mathscr{L}}(D, G)$ has minimum distance $d = n - k + 1$ if and only if

$$N(k, P, D) = 0\ .$$

And the minimum distance $d = n - k$ if and only if

$$N(k, P, D) > 0\ .$$

Proof: We have already seen that the minimum distance of $C_{\mathscr{L}}(D, G)$ has two choices: $n-k$, $n-k+1$. So $C_{\mathscr{L}}(D, G)$ is not MDS, i.e., $d = n - k$ if and only if there is a function $f \in \mathscr{L}(G)$ such that the evaluation $ev(f)$ has weight $n - k$. This is equivalent to that f has k zeros in D, say $P_{i_1}, \cdots, P_{i_k}$. That is

$$\mathrm{div}(f) \geq -(k-1)O - P + (P_{i_1} + \cdots + P_{i_k}),$$

which is equivalent to

$$\mathrm{div}(f) = -(k-1)O - P + (P_{i_1} + \cdots + P_{i_k}).$$

The existence of such an f is equivalent to saying

$$P_{i_1} \oplus \cdots \oplus P_{i_k} = P.$$

Namely, $N(k, P, D) > 0$. It follows that the AG code $C_{\mathscr{L}}(D, G)$ has minimum distance $n - k + 1$ if and only if $N(k, P, D) = 0$. ∎

Proposition 4.1 establishes the relation between minimum distance of ECAG code and SSP on the rational point group of the elliptic curve. Together with Corollaries 3.4 and 3.5, we obtain the main result of this paper, Theorem 1.3.

Acknowledgment

This paper was written when the first author was visiting the Department of Mathematics, University of Delaware and the Department of Mathematics, University of California, Irvine. The first author warmly thanks both departments for their hospitality and Professor Qing Xiang for his help. The third author would like to thank Maosheng Xiong for inviting him to visit Hong Kong University of Science and Technology during ISIT 2015.

The authors would like to thank the anonymous reviewers for reminding the references [8] and [9].

The work of Jiyou Li is supported by the Ky and Yu-Fen Fan Fund Travel Grant from the AMS, the National Science Foundation of China (11001170) and the National Science Foundation of Shanghai Municipal (13ZR1422500). The research of Daqing Wan is partially supported by NSF. This research of Jun Zhang is supported by the National Key Basic Research Program of China (2013CB834204), the National Natural Science Foundation of China (61171082, 10990011 and 60872025).

References

[1] A. Vardy, "The intractability of computing the minimum distance of a code," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1757–1766, 1997.

[2] Q. Cheng, "Hard problems of algebraic geometry codes," *IEEE Transactions on Information Theory*, vol. 54, pp. 402–406, 2008.

[3] J. Li and D. Wan, "A new sieve for distinct coordinate counting," *Science China-mathematics*, vol. 53, pp. 1–12, 2010.

[4] I. Dumer, D. Micciancio, and M. Sudan, "Hardness of approximating the minimum distance of a linear code," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 22–37, 2003.

[5] Q. Cheng and D. Wan, "A deterministic reduction for the gap minimum distance problem:[extended abstract]," in *Proceedings of the 41st annual ACM symposium on Theory of computing*. ACM, 2009, pp. 33–38.

[6] J. Zhang, F.-W. Fu, and D. Wan, "Stopping sets of algebraic geometry codes," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1488–1495, March 2014.

[7] J. Li and D. Wan, "Counting subset sums of finite abelian groups," *Journal of Combinatorial Theory, Series A*, vol. 119, no. 1, pp. 170 – 182, 2012.

[8] C. Munuera, "On the main conjecture on geometric MDS codes," *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1573–1577, Sep 1992.

[9] S. Ball, "On sets of vectors of a finite vector space in which every subset of basis size is a basis," *Journal of the European Mathematical Society*, vol. 14, no. 3, pp. 733–748, 2012.

[10] H. Stichtenoth, *Algebraic function fields and codes*, 2nd ed., ser. Graduate Texts in Mathematics. Berlin: Springer-Verlag, 2009, vol. 254.

[11] J. Silverman, *The arithmetic of elliptic curves*, 2nd ed., ser. Graduate Texts in Mathematics. Dordrecht: Springer, 2009, vol. 106.