COMBINATORIAL PROBLEMS IN FINITE FIELDS AND SIDON SETS

JAVIER CILLERUELO

Abstract. We use Sidon sets to present an elementary method to study some combinatorial problems in finite fields, such as sum product estimates, solvability of some equations and the distribution of their solutions. We obtain classic and more recent results avoiding the use of exponential sums, the usual tool to deal with these problems.

1. Introduction

The aim of the present work is to introduce a new elementary method to study a class of combinatorial problems in finite fields: incidence problems, sum-product estimates, solvability of some equations, distribution of solutions of exponential equations, etc.

The main tool in our approach are Sidon sets, which are important objects in combinatorial number theory.

In Section §2 we present Theorem 2.1, which is the main tool in our method. To illustrate how this method works, we include in this section two easy applications of this theorem. The first one recovers a result of Vinh [15] about the number of incidences between P points and L lines in the field $\mathbb{F}_q$. The second one proves that if $A = \{g^x : \ 0 \le x \le (\sqrt{2} + o(1))q^{3/4}\}$ and g is a generator of $\mathbb{F}_q^*$, then $A - A = \mathbb{F}_q$. This improves and generalizes previous results obtained in the prime fields $\mathbb{F}_p$ by Garaev-Kueh [4], Konyagin [9] and García [6].

Section §3 is devoted to sum-product estimates. Garaev [2] used character sums to give the nontrivial lower estimate $\max(|A+A|, |AA|) \gg \min(\sqrt{|A|p}, |A|^2/\sqrt{p})$ in $\mathbb{F}_p$. Lemma 3.1, which is an easy consequence of Theorem 2.1, gives a nontrivial upper bound for the number of elements of a dense Sidon set in an arbitrary set B when $|B + B|$ is small. We use this upper bound to give a quick proof of Garaev's estimate and related results.

Sárközy [11, 12] proved the solvability of the equations $x_1x_2 + x_3x_4 = 1$ and $x_1x_2 = x_3 + x_4,\ x_i \in A_i$ for arbitrary sets $A_i \subset \mathbb{F}_p$ when $|A_1||A_2||A_3||A_4| \gg p^3$. This result was extended to any field $\mathbb{F}_q$ in [7]. The proof is based in estimates of exponential sums and they asked for an elementary algebraic proof of the solvability of these equations (problem 3 of [1]). Our method provides a proof of this kind. Actually, Sarközy's equations are special cases of more general equations which we study in section §4.

In section §5, we apply our method to study the distribution of the solutions of some equations in $\mathbb{F}_q$. As an example we prove that if g is a generator of $\mathbb{F}_q^*$, then for any intervals $I, J \subset \mathbb{Z}_{q-1}$ we have

$$|\{(x,y) \in I \times J :\ g^x - g^y = 1\}| = \frac{|I||J|}{q} + O\left(q^{1/2} e^{c\sqrt{\log(|I||J|/q^{3/2}+1)}}\right) q,$$

for some positive constant c. The error term is smaller than the error term obtained by Garaev [3].

2. Sidon sets

Let G be a finite abelian group. For any sets $A, B \subset G$ and $x \in G$, we write $r_{A-B}(x)$ for the number of representations of $x = a - b,\ a \in A,\ b \in B$. We have the well known identities

$$\sum_{x \in G} r_{A-B}(x) = |A||B|, \tag{2.1}$$

$$\sum_{x \in G} r^2_{A-B}(x) = \sum_{x \in G} r_{A-A}(x) r_{B-B}(x). \tag{2.2}$$

Definition 1. *We say that a set $\mathcal{A} \subset G$ is a Sidon set if $r_{\mathcal{A}-\mathcal{A}}(x) \le 1$ whenever $x \neq 0$.*

By counting the number of differences $a - a'$, we can see that if $\mathcal{A}$ is a Sidon set, then $|\mathcal{A}| < \sqrt{|G|} + 1/2$. The most interesting Sidon sets are those which have large cardinality, that is, $|\mathcal{A}| = \sqrt{|G|} - \delta$ where $|\delta|$ is a small number. We state our main theorem.

Theorem 2.1. *Let $\mathcal{A}$ be a Sidon set in a finite abelian group G with $|\mathcal{A}| = \sqrt{|G|} - \delta$. Then, for all $B, B' \subset G$ we have*

$$|\{(b,b') \in B \times B',\ b + b' \in \mathcal{A}\}| = \frac{|\mathcal{A}|}{|G|}|B||B'| + \theta(|B||B'|)^{1/2}|G|^{1/4},$$

with $|\theta| < 1 + \frac{|B|}{|G|}\max(0,\delta)$.

Proof. Since $\mathcal{A}$ is a Sidon set,

$$\begin{aligned}\sum_{x\in G} r_{B-B}(x)r_{\mathcal{A}-\mathcal{A}}(x) &= |\mathcal{A}||B| + \sum_{x\neq 0} r_{B-B}(x)r_{\mathcal{A}-\mathcal{A}}(x)\\ &\leq |\mathcal{A}||B| + \sum_{x\neq 0} r_{B-B}(x) = |\mathcal{A}||B| + |B|^2 - |B|.\end{aligned}$$

Using this inequality and identities (2.1) and (2.2) we have

$$\begin{aligned}\sum_{x\in G}\left(r_{\mathcal{A}-B}(x) - \frac{|\mathcal{A}||B|}{|G|}\right)^2 &= \sum_{x\in G} r_{B-B}(x)r_{\mathcal{A}-\mathcal{A}}(x) - \frac{|\mathcal{A}|^2|B|^2}{|G|}\\ &\leq |B|(|\mathcal{A}|-1) + |B|^2\frac{|G|-|\mathcal{A}|^2}{|G|}.\end{aligned} \tag{2.3}$$

We observe that

$$|\{(b,b')\in B\times B',\ b+b'\in\mathcal{A}\}| - \frac{|B||B'||\mathcal{A}|}{|G|} = \sum_{b'\in B'}\left(r_{\mathcal{A}-B}(b') - \frac{|\mathcal{A}||B|}{|G|}\right).$$

Applying the Cauchy-Schwarz inequality, taking (2.3) and $|\mathcal{A}| = |G|^{1/2} - \delta$ into account we obtain

$$\begin{aligned}\left|\sum_{b'\in B'}\left(r_{\mathcal{A}-B}(b') - \frac{|\mathcal{A}||B|}{|G|}\right)\right|^2 &\leq |B'|\left(|B|(|\mathcal{A}|-1) + |B|^2\frac{|G|-|\mathcal{A}|^2}{|G|}\right)\\ &= |B'||B|\left(|G|^{1/2} - \delta - 1 + |B|\frac{\delta(2|G|^{1/2}-\delta)}{|G|}\right)\\ &< |B||B'||G|^{1/2}\left(1 + 2\max(0,\delta)\frac{|B|}{|G|}\right).\end{aligned}$$

□

The Sidon sets we will consider in applications satisfy $\delta \leq 1$ and $|B| = o(|G|)$. In these cases we have $|\theta| \leq 1 + o(1)$.

2.1. **Examples of dense Sidon sets.** The three families of Sidon sets we will describe next, have maximal cardinality in their ambient group G. Let g be a generator of $\mathbb{F}_q^*$.

Example 1. *Let $p(x), r(x) \in \mathbb{F}_q[X]$ be polynomials of degree ≤ 2 such that $p(x) - \mu r(x)$ is not a constant for any $\mu \in \mathbb{F}_q$. The set*

$$\mathcal{A} = \{(p(x), r(x)) : \ x \in \mathbb{F}_q\}$$

is a Sidon set in $\mathbb{F}_q \times \mathbb{F}_q$. In particular, the set $\mathcal{A} = \{(x, x^2) : \ x \in \mathbb{F}_q\}$ is a Sidon set.

We have to check that when $(e_1, e_2) \neq (0,0)$ the relation $(p(x_1), r(x_1)) - (p(x_2), r(x_2)) = (e_1, e_2)$ uniquely determines x_1 and x_2. If $p(x)$ is linear then from $p(x_1) - p(x_2) = e_1$ we obtain $x_1 = x_2 + \lambda$ for some λ. Thus, $r(x_2 + \lambda) - r(x_2) = e_2$ is a linear equation and we obtain x_2 and then x_1. If $p(x)$ is quadratic we consider μ such that $p(x) - \mu r(x)$ is a linear polynomial and we proceed as above.

Example 2. *For any generator g of $\mathbb{F}_q^*$, the set*

$$\mathcal{A} = \{(x, g^x) : \ x \in \mathbb{Z}_{q-1}\} \tag{2.4}$$

is a Sidon set in $\mathbb{Z}_{q-1} \times \mathbb{F}_q$.

Sometimes we will describe this set as $\mathcal{A} = \{(\log x, x) : \ x \in \mathbb{F}_q^*\}$ where $\log x = \log_g x$ is the discrete logarithm.

From $(x_1, g^{x_1}) - (x_2, g^{x_2}) = (e_1, e_2) \neq (0,0)$ we have $x_1 - x_2 \equiv e_1 \pmod{q-1}$ and hence $g^{x_1} = g^{e_1 + x_2}$. Putting this in $g^{x_1} - g^{x_2} = e_2$ we get $g^{x_2}(g^{e_1} - 1) = e_2$.

If $e_1 = 0$ then $e_2 = 0$, but we have assumed that $(e_1, e_2) \neq (0,0)$. If $e_1 \neq 0$ the last equality determines x_2, and then x_1.

Example 3. *For any pair of generators g_1, g_2 of $\mathbb{F}_q^*$, the set*

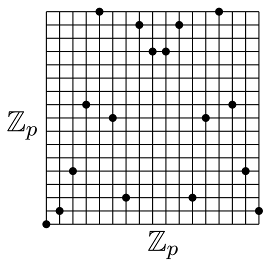
$$\mathcal{A} = \{(x, y) \in \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1} : \ g_1^x + g_2^y = 1\} \tag{2.5}$$

is a Sidon set in $\mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}$. Since translations preserve Sidoness property, for any $\lambda \neq 0$, the sets $\mathcal{A} = \{(x, y) : \ g_1^x + g_2^y = \lambda\}$ and $\mathcal{A} = \{(x, y) : \ g_1^x - g_2^y = \lambda\}$ are also Sidon sets.

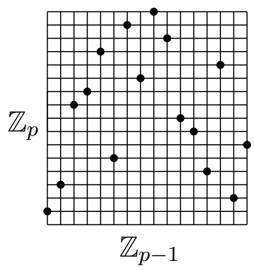
To see that $\mathcal{A}$ is a Sidon set we have to prove that if $(e_1, e_2) \neq (0,0)$ then the equation $(x_1, y_1) - (x_2, y_2) = (e_1, e_2)$ determines x_1, x_2 under the conditions $g_1^{x_1} + g_2^{y_1} = g_1^{x_2} + g_2^{y_2} = 1$ in $\mathbb{F}_q$. We observe that $x_1 - x_2 \equiv e_1 \pmod{(q-1)}$ and $y_1 - y_2 \equiv e_2 \pmod{(q-1)}$ imply that $g_1^{x_1} = g_1^{x_2 + e_1}$ and $g_2^{y_1} = g_2^{y_2 + e_2}$ in $\mathbb{F}_q$ and we obtain $g_1^{x_2 + e_1} + g_2^{y_2 + e_2} = g_1^{x_2} + g_2^{y_2} = 1$ in $\mathbb{F}_q$. Thus $g_2^{y_2}(g_2^{e_2} - g_1^{e_1}) = 1 - g_1^{e_1}$. If

$e_1 \neq 0$ and $g_2^{e_2} \neq g_1^{e_1}$ we obtain y_2 and then x_2, x_1 and y_1. If $e_1 = 0$ or $g_2^{e_2} = g_1^{e_1}$, the equation has not solutions unless $(e_1, e_2) = (0, 0)$.

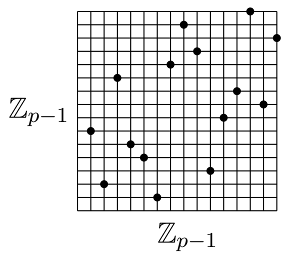
When $q = p$ is a prime number we can identify $\mathbb{F}_p$ with $\mathbb{Z}_p$. We show in the pictures below the three examples of Sidon sets described above when $p = 17$ respectively.



Example 1



Example 2



Example 3

The Sidon sets given in examples, with $q, q-1$ and $q-2$ elements respectively, have maximal cardinality in their ambient groups. The values of $\delta = |G|^{1/2} - |\mathcal{A}|$ are $\delta = 0, 1/2 - o(1)$ and 1 respectively. We finish this section with two easy applications of Theorem 2.1.

2.2. **Incidence of lines and points in $\mathbb{F}_q \times \mathbb{F}_q$.** Let $I(P, L) = |\{(p, l) \in P \times L : p \in L\}|$ be the number of incidences between a set P of points and set L of lines in $\mathbb{F}_q \times \mathbb{F}_q$, that is

$$I(P, L) = |\{(p, l) \in P \times L : \ p \in l\}| \,.$$

Vinh [15] proved that $I(P, L) \leq \frac{|P||L|}{q} + q^{1/2}\sqrt{|P||L|}$. We get an asymptotic estimate for $I(P, L)$ as a straightforward consequence of Theorem 2.1.

Theorem 2.2. *Let L be a set of lines and let P be a set of points in $\mathbb{F}_q \times \mathbb{F}_q$. The following asymptotic formula holds:*

$$I(P, L) = \frac{|P||L|}{q} + O(q^{1/2}\sqrt{|P||L|}). \tag{2.6}$$

Proof. Let $L = \{y = \lambda_i x + \mu_i : \ 1 \leq i \leq |L|\}$ and $P = \{(p_j, q_j) : \ 1 \leq j \leq |P|\}$. We consider the set $\mathcal{A} = \{(\log x, x)\}$ described in Example (2) and the sets

$$B = \{(\log \lambda_i, -\mu_i) : \ 1 \leq i \leq |L|\}, \qquad B' = \{(\log p_j, q_j) : \ 1 \leq j \leq |P|\}.$$

We observe that each incidence corresponds to a solution of $\lambda_i p_j = q_j - \mu_i$ and the number of solutions of this equation is $|\{(b,b') \in B \times B' :\ b + b' \in \mathcal{A}\}|$. The result follows in view of Theorem 2.1. $\square$

2.3. **The difference set** $\{g^x - g^y :\ 0, \le x, y \le L\}$**.** Let g be a generator of $\mathbb{F}_q^*$. Many authors have studied the problem of determining the smallest number M such that $\{g^x - g^y :\ 0 \le x, y \le M\} = \mathbb{F}_q$ in the case q is a prime p.

From the result of Rudnick and Zaharescu [10] it follows that one can take any integer $M \ge c_0 p^{3/4} \log p$ where c_0 is a suitable constant. This range has been improved to $M > cp^{3/4}$ by Garaev and Kueh [4] and independently by Konyagin [9]. The best known admissible value for the constant c has been $c = 2^{5/4}$ due to García [6]. Our approach improves this further to the following statement.

Theorem 2.3. *Let g be a generator of $\mathbb{F}_q^*$. For any $\epsilon > 0$ and $q > q(\epsilon)$ we have*

$$\left\{g^x - g^y :\ 0 \le x, y < (\sqrt{2} + \epsilon) q^{3/4}\right\} = \mathbb{F}_q.$$

Proof. Suppose $\lambda \notin \{g^x - g^y :\ 0 \le x, y \le L\}$ and consider in $G = \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}$ the Sidon set (see Example 3) $\mathcal{A} = \{(x,y) :\ g^x - g^y = \lambda\}$. We observe that

$$(x,y) \in \mathcal{A} \iff (y,x) + (\frac{q-1}{2}, \frac{q-1}{2}) \in \mathcal{A}.$$

Thus, it is clear that if $b, b' \in B = [0, L/2]^2 + \left\{(0,0), \left(\frac{q-1}{2}, \frac{q-1}{2}\right)\right\}$ then $b + b' \notin \mathcal{A}$. In other words, $|\{(b,b') \in B \times B,\ b + b' \in \mathcal{A}\}| = 0$.

On the other hand, Theorem 2.1 implies that

$$0 = |\{(b,b') \in B \times B,\ b + b' \in \mathcal{A}\}| \ge \frac{|\mathcal{A}||B|^2}{|G|} - \Big(1 + \frac{|B|}{|G|}\Big)|B||G|^{1/4}.$$

Thus

$$|B| \le \frac{|G|^{5/4}}{|\mathcal{A}| - |G|^{1/4}} = \frac{(q-1)^{5/2}}{q - 2 - (q-1)^{1/2}} < q^{3/2}(1 + o(1))$$

and the theorem follows in view of $|B| = 2(1 + [L/2])^2$. $\square$

3. Sum-product estimates

We will deduce some sum-product estimates form the following lemma.

Lemma 3.1. *Let $\mathcal{A}$ be a Sidon set in G with $|\mathcal{A}| = |G|^{1/2} - \delta$. For any subsets $B, B' \subset G$ we have*

$$|\mathcal{A} \cap B| \leq \frac{|B+B'||\mathcal{A}|}{|G|} + \theta \left(\frac{|B+B'|}{|B'|} \right)^{1/2} |G|^{1/4},$$

for some θ with $|\theta| \leq 1 + \max(0, \delta)\frac{|B'|}{|G|}$.

Proof. Indeed, by Theorem 2.1,

$$\begin{aligned} |B'||\mathcal{A} \cap B| &= |\{(-b', b+b') : \ b \in B, \quad b' \in B', \ -b' + (b+b') \in \mathcal{A}\}| \\ &\leq |\{(b', b'') : \ b' \in (-B') \times (B+B'), \ b' + b'' \in \mathcal{A}\}| \\ &\leq \frac{|\mathcal{A}||B'||B+B'|}{|G|} + \theta \sqrt{|B'||B+B'|} |G|^{1/4} \end{aligned}$$

for some θ with $|\theta| \leq 1 + \max(0, \delta)\frac{|B'|}{|G|}$. The lemma follows. □

Theorem 3.1 (Garaev [2])**.** *Let $A_1, A_2 \subset \mathbb{F}_q^*$ and $A_3 \subset \mathbb{F}_q$. We have*

$$\max(|A_1 A_2|, |A_1 + A_3|) \gg \min\left(\sqrt{|A_1| q}, \sqrt{|A_1|^2 |A_2||A_3|/q} \right). \tag{3.1}$$

Proof. We consider the Sidon set $\mathcal{A} = \{(\log x, x) : \ x \in \mathbb{F}_q^*\}$ described in example 2 and the sets $B = (\log A_1) \times A_1$ and $B' = (\log A_2) \times A_3$. Since all the elements $(\log a_1, a_1)$ are in $\mathcal{A}$ we have that $|\mathcal{A} \cap B| = |A_1|$. On the other hand we observe that $|B + B'| = |A_1 A_2||A_1 + A_3|$. Lemma 3.1 implies the inequality

$$|A_1| \leq \frac{|A_1 A_2||A_1 + A_3|}{q} + \theta \sqrt{q \frac{|A_1 A_1||A_1 + A_3|}{|A_2||A_3|}},$$

for some θ with $|\theta| \leq 1$, which in turn implies (3.1). □

We can mimic this proof to get the following sum-product estimates.

Theorem 3.2 (Garaev-Shen [5])**.** *Let $A_1, A_2, A_3 \subset \mathbb{F}_q^*$. We have*

$$\max(|(A_1 + 1)A_2|, |A_1 A_3|) \gg \min\left(\sqrt{|A_1| q}, \sqrt{|A_1|^2 |A_2||A_3|/q} \right).$$

Proof. We consider the Sidon set $\mathcal{A} = \{(x, y) : \ g^x - g^y = 1\}$, the sets $B = \log(A_1 + 1) \times \log A_1$ and $B' = \log A_2 \times \log A_3$ and proceed as in the proof of Theorem 3.1. □

Theorem 3.3 (Solymosi [14], Hart-Li-Shen [8]). *Let $p(x), q(x) \in \mathbb{F}_q[X]$ be polynomials of degree ≤ 2 such that $p(x) - \mu q(x)$ is not a constant for any $\mu \in \mathbb{F}_q$. For any $A_1, A_2, A_3 \subset \mathbb{F}_q$ we have*

$$\max(|p(A_1) + A_2|, |q(A_1) + A_3|) \gg \min\left(\sqrt{|A_1|q}, \sqrt{|A_1|^2|A_2||A_3|/q}\right).$$

Proof. We consider the Sidon set $\mathcal{A} = \{(p(x), q(x)) : x \in \mathbb{F}_q\}$, the sets $B = p(A_1) \times q(A_1)$ and $B' = A_2 \times A_3$ and proceed as in the proof of Theorem 3.1. $\square$

Solymosi [14] proved that if $\{(x, f(x)) : x \in \mathbb{F}_q\} \subset \mathbb{F}_q \times \mathbb{F}_q$ is a Sidon set then $\max(|A + A|, |f(A) + f(A)|) \gg \min(\sqrt{|A|q}, |A|^2/\sqrt{q})$.

4. Equations in $\mathbb{F}_q$

We start with the easiest example which, however, we have not seen in the literature.

Theorem 4.1. *For any $x \in \mathbb{F}_q$, let $X(x), Y(x)$ be any pair of subsets of $\mathbb{F}_q$ and put $T = \left(\sum_x |X(x)|\right)\left(\sum_x |Y(x)|\right)$. Then, the number of solutions S of*

$$x' + y' = (x + y)^2, \quad x' \in X(x),\ y' \in Y(y)$$

is

$$S = \frac{T}{q} + \theta\sqrt{qT}$$

for some θ with $|\theta| \le 1$.

Proof. We consider the Sidon set $\mathcal{A} = \{(x, x^2) : x \in \mathbb{F}_q\}$ and the sets

$$B = \{(x, x') : x' \in X(x)\}, \qquad B' = \{(y, y') : y' \in Y(y)\}.$$

From the definition, $(x, x') + (y, y') \in \mathcal{A} \iff x' + y' = (x + y)^2$. Thus $S = |\{(b, b') \in B \times B' : b + b' \in \mathcal{A}\}|$ and we apply Theorem 2.1. $\square$

Corollary 4.1. *Let $A_1, A_2, A_3, A_4 \subset \mathbb{F}_q$. Then, the number of solutions of the equation*

$$x_1 + x_2 = (x_3 + x_4)^2, \qquad x_i \in A_i \tag{4.1}$$

is

$$S = \frac{|A_1||A_2||A_3||A_4|}{q} + \theta\sqrt{q|A_1||A_2||A_3||A_4|},$$

for some θ with $|\theta| \le 1$. In particular, the number of solutions of

$$x_1 + x_2 = z^2, \quad x_1 \in A_1,\ x_2 \in A_2,\ z \in \mathbb{F}_q \tag{4.2}$$

is

$$|A_1||A_2| + \theta\sqrt{|A_1||A_2|q},$$

for some θ with $|\theta| \le 1$.

Proof. The first part of the statement follows from Theorem 4.1 by taking

$$X(x) = \begin{cases} A_1, & x \in A_3 \\ \emptyset & \text{otherwise} \end{cases} \qquad \text{and} \qquad Y(x) = \begin{cases} A_2, & x \in A_4 \\ \emptyset & \text{otherwise.} \end{cases}$$

The second part of the statement follows from the fact that if $A_3 = A_4 = \mathbb{F}_q$ then each solution of (4.2) corresponds to exactly q solutions of (4.1). □

Shkredov [13] used Weil's bound for exponential sums with multiplicative characters to prove the following result for $q = p$ prime and the condition $|X_1||X_2| > 20p$.

Corollary 4.2. *Let $X_1, X_2 \subset \mathbb{F}_q$, $|X_1||X_2| > 2q$. Then there exist $x, y \in \mathbb{F}_q$ such that $x + y \in X_1$ and $xy \in X_2$.*

Proof. The number of such that pairs (x, y) is equal to the number of solutions of the equation

$$(x_1/2 - z)(x_1/2 + z) = x_2, \qquad x_1 \in X_1,\ x_2 \in X_2,\ z \in \mathbb{F}_q.$$

We observe that this equation is equivalent to the equation $(x_1/2)^2 - x_2 = z^2$. The second part of Corollary 4.1 applied to the sets $A_1 = \{x_1^2/2 :\ x_i \in X_1\}$ and $A_2 = -X_2$ proves that the number of solutions of the equation $(x_1/2)^2 - x_2 = z^2, \quad x_1 \in X_1,\ x_2 \in X_2,\ z \in \mathbb{F}_q$ is at least $|A_1||A_2| - q^{1/2}\sqrt{|A_1||A_2|}$. This quantity is positive when $|A_1||A_2| > q$ and the theorem follows since $|A_1| \ge |X_1|/2$ and $|A_2| = |X_2|$. □

Sárközy [11, 12] using exponential sums, obtained asymptotic formula for the number of solutions of the congruences $x_1x_2 - x_3x_4 \equiv \lambda \pmod p$ and $x_1x_2 - x_3 - x_4 \equiv \lambda \pmod p$, $x_i \in X_i$. In [7] these results have been proved in any finite fields. We derive Sarkőzy's results directly from our Theorem 2.1.

Theorem 4.2. *For $x \in \mathbb{F}_q^*$, $y \in \mathbb{F}_q^*$ let $X(x), Y(y)$ be any subsets of $\mathbb{F}_q$. Then for the number S of solutions of the equation*

$$x' + y' = xy, \quad x' \in X(x),\ y' \in Y(y)$$

we have

$$S = \frac{T}{q} + \theta\sqrt{qT},$$

for some θ with $|\theta| \le 1 + o(1)$, where $T = \left(\sum_x |X(x)|\right)\left(\sum_x |Y(x)|\right)$.

Proof. We consider the Sidon set $\mathcal{A} = \{(x, g^x) :\ x \in \mathbb{Z}_{q-1}\}$ and the sets

$$B = \{(\log x, x') :\ x' \in X(x)\}, \qquad B' = \{(\log y, y') :\ y' \in Y(y)\}.$$

We observe that

$$(\log x, x') + (\log y, y') \in \mathcal{A} \iff x' + y' = g^{\log x + \log y} = xy.$$

Thus $S = |\{(b, b') \in B \times B' :\ b + b' \in \mathcal{A}\}|$ and then we apply Theorem 2.1. $\square$

Corollary 4.3. *Let $X_1, X_2 \subset \mathbb{F}_q^*$ and $X_3, X_4 \subset \mathbb{F}_q$. The number S of solutions of the equation*

$$x_1x_2 = x_3 + x_4, \quad x_i \in X_i,$$

is

$$S = \frac{|X_1||X_2||X_3||X_4|}{q} + \theta\sqrt{|X_1||X_2||X_3||X_4|q},$$

for some θ with $|\theta| \le 1 + o(1)$.

Proof. We take $X(x)$ and $Y(y)$ as in Corollary 4.1 and use Theorem 4.2. $\square$

Corollary 4.4. *Let $X_1, X_2 \subset \mathbb{F}_q^*$ and $X_3, X_4 \subset \mathbb{F}_q$. The number S of solutions of the equation*

$$x_2x_3 - x_1x_4 = 1, \quad x_i \in X_i,$$

is

$$S = \frac{|X_1||X_2||X_3||X_4|}{q} + \theta\sqrt{|X_1||X_2||X_3||X_4|q}, \qquad |\theta| \le 1 + o(1).$$

Proof. In Theorem 4.2 we take

$$X(x) = \begin{cases} xX_3, & x \in X_1^{-1} \\ \emptyset & \text{otherwise} \end{cases} \qquad \text{and} \qquad Y(y) = \begin{cases} -yX_4, & y \in X_2^{-1} \\ \emptyset & \text{otherwise} \end{cases}.$$

In this way we arrive at the equation $x_1^{-1}x_2^{-1} = x_1^{-1}x_3 - x_2^{-1}x_4$, which is equivalent to the equation of the corollary. $\square$

Theorem 4.3. *For $x \in \mathbb{F}_q^*,\ y \in \mathbb{F}_q^*$, let $X(x), Y(y)$ be any subsets of $\mathbb{F}_q^*$. The number S of solutions of the equation*

$$xy - x'y' = 1, \qquad x, y \in \mathbb{F}_q^*,\ x' \in X(x),\ y' \in Y(y),$$

is

$$S = \frac{T}{q} + \theta\sqrt{Tq},$$

for some θ with $|\theta| \le 1 + o(1)$, where $T = \Big(\sum_x |X(x)|\Big)\Big(\sum_x |Y(x)|\Big)$.

Proof. We consider the Sidon set $\mathcal{A} = \{(x, y) : g^x - g^y = 1\} \subset \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}$ and the sets $B = \{(\log x, \log x') :\ x' \in X(x)\}$ and $B' = \{(\log y, \log y') :\ y' \in Y(y)\}$. It is clear that $S = |\{(b, b') \in B \times B' :\ b + b' \in \mathcal{A}\}|$. Now we apply Theorem 2.1. $\square$

We observe that this theorem also gives an alternative proof of Corollary 4.4 by taking $X(x)$ and $Y(y)$ as in Corollary 4.1.

5. Distribution of Sidon sets and applications

Let $\mathcal{A}$ be a Sidon set in G. For any set $B \subset G$ we denote by $D_{\mathcal{A}}(B)$ the discrepancy $D_{\mathcal{A}}(B) = \left| |\mathcal{A} \cap B| - \frac{|B||\mathcal{A}|}{|G|} \right|$.

The following lemma and Lemma 3.1 will be the main tools to prove asymptotic estimates for $|\mathcal{A} \cap B|$ in a class of problems. For simplicity we restrict ourselves to the cases when $\mathcal{A}$ is one of the three Sidon sets described in Section 2.

Lemma 5.1. *Let $\mathcal{A}$ be one of the three Sidon sets described in section 2 and $B \subset G$. For any set $C \subset G$, there exists $c \in C$ such that*

$$D_{\mathcal{A}}(B) \le 2\Big(q\frac{|B|}{|C|}\Big)^{1/2} + D_{\mathcal{A}}(B^c) + D_{\mathcal{A}}(B_c)$$

where $B^c = B \setminus (B + c)$ and $B_c = (B + c) \setminus B$.

Proof. We can write

(5.1)
$$D_{\mathcal{A}}(B)=\left|\frac{1}{|C|}\sum_{c\in C}\Big(|\mathcal{A}\cap(B+c)|-\frac{|\mathcal{A}||B|}{|G|}\Big)+\frac{1}{|C|}\sum_{c\in C}\Big(|\mathcal{A}\cap B|-|\mathcal{A}\cap(B+c)|\Big)\right|.$$

We observe that for the first sum we have
$$\sum_{c\in C}\Big(|\mathcal{A}\cap(B+c)|-\frac{|\mathcal{A}||B|}{|G|}\Big)=|\{(b,c)\in B\times C:\ b+c\in\mathcal{A}\}|-\frac{|\mathcal{A}||B||C|}{|G|}.$$

Hence, by Theorem 2.1, the absolute value of this sum is bounded by $2\Big(q|B||C|\Big)^{1/2}$.

For the second sum in (5.1), it follows from $|B_c|=|B^c|$ that
$$\begin{aligned}|\mathcal{A}\cap B|-|\mathcal{A}\cap(B+c)|&=|\mathcal{A}\cap B_c|-|\mathcal{A}\cap B^c|\\&=\left(|\mathcal{A}\cap B_c|-\frac{|\mathcal{A}||B_c|}{|G|}\right)-\left(|\mathcal{A}\cap B^c|-\frac{|\mathcal{A}||B^c|}{|G|}\right).\end{aligned}$$

Thus
$$\begin{aligned}\left|\frac{1}{|C|}\sum_{c\in C}\Big(|\mathcal{A}\cap B|-|\mathcal{A}\cap(B+c)|\Big)\right|&\le\frac{1}{|C|}\sum_{c\in C}\Big(D_{\mathcal{A}}(B_c)+D_{\mathcal{A}}(B^c)\Big)\\&\le\max_{c\in C}\Big(D_{\mathcal{A}}(B_c)+D_{\mathcal{A}}(B^c)\Big).\end{aligned}$$

□

In the special case when B is a subgroup we can take $C=B$ and then $B^c=B_c=\emptyset$ for any $c\in C$. Thus, in this case we have
$$D_{\mathcal{A}}(B)\ll q^{1/2}.$$

As a corollary we obtain a well known result on the Fermat equation in finite fields.

Corollary 5.1. *Let Q,Q' be subgroups of $\mathbb{F}_q^*$. We have*
$$|\{(x,y)\in Q\times Q':\ x+y=1\}|=\frac{|Q||Q'|}{q}+O(\sqrt{q}).$$
In particular, if $p\gg(rs)^2$ the Fermat congruence $x^r+y^s\equiv 1 \pmod p$ has nontrivial solutions.

Proof. Consider the Sidon set $\mathcal{A}=\{(x,y):\ g^x+g^y=1\}$ and take $B=C=Q\times Q'$. □

In applications, the strategy is to take a large set C such that $|B^c|$ and $|B_c|$ are small compared with $|B|$. This is possible when B has some specific regularity properties (subgroups, cartesian product of arithmetic progressions, convex sets, etc.) We illustrate our method with an example.

Theorem 5.1. *For any intervals $I, J \subset \mathbb{Z}_{q-1}$ and g a generator of $\mathbb{F}_q^*$ we have*

$$|\{(x,y) \in I \times J :\ g^x - g^y = 1\}| = \frac{|I||J|}{q} + O\left(q^{1/2} e^{c\sqrt{\log(|I||J|/q^{3/2}+1)}}\right),$$

for some constant $c > 0$.

Proof. Indeed we will prove that for any positive r we have

$$|\{(x,y) \in I \times J :\ g^x - g^y = 1\}| = \frac{|I||J|}{q} + \theta\, 6^r \Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/r} + 1\Big)\sqrt{q}, \tag{5.2}$$

for some θ with $|\theta| \le 1$. The choice $r = \lceil \sqrt{\log(|I||J|/q^{3/2}+1)} \rceil$ gives the error term in Theorem 5.1. We proceed by induction on r. We consider the Sidon set $\mathcal{A} = \{(x,y) :\ g^x - g^y = 1\} \subset \mathbb{Z}_{q-1} \times \mathbb{Z}_{q-1}$ and the set $B = I \times J$. Then, applying Lemma 3.1, we get

$$|\mathcal{A} \cap B| \le \frac{|B+B|}{q} + 2\sqrt{q\frac{|B+B|}{|B|}} \le \frac{4|I||J|}{q} + 4\sqrt{q}.$$

Since $D_{\mathcal{A}}(B) \le \max(\frac{|B||\mathcal{A}|}{|G|}, |\mathcal{A} \cap B|)$, we have that $D_{\mathcal{A}}(B) \le \frac{4|I||J|}{q} + 4\sqrt{q}$, which proves (5.2) for $r = 1$. Now we assume that (5.2) is true for some r and we proved it for $r+1$.

We consider the auxiliar set $C = I' \times J'$ where $I' = \{0, \dots, \lfloor \alpha|I| \rfloor\}$ and $J' = \{0, \dots, \lfloor \alpha|J| \rfloor\}$ for a suitable $\alpha \le 1$. We observe that $|C| \ge \alpha^2|I||J|$. Lemma 5.1 gives $D_{\mathcal{A}}(B) \le 2\frac{q^{1/2}}{\alpha} + D_{\mathcal{A}}(B^c) + D_{\mathcal{A}}(B_c)$ for some $c \in C$ where $B^c = B \setminus (B+c)$ and $B_c = (B+c) \setminus B$. Now we observe that $B+c$ is a translation of the rectangle $B = I \times J$. Thus we can write $B^c = B_1 \cup B_2$ and $B_c = B_3 \cup B_4$ where the sets B_i are rectangles with $|B_i| \le \alpha|I||J|$.

Thus, $D_{\mathcal{A}}(B) \le 2\frac{\sqrt{q}}{\alpha} + D_{\mathcal{A}}(B_1) + D_{\mathcal{A}}(B_2) + D_{\mathcal{A}}(B_3) + D_{\mathcal{A}}(B_4)$.

Taking into account the induction hypothesis for each B_i we have

$$D_{\mathcal{A}}(B) \le \left(\frac{2}{\alpha} + 4 \cdot 6^r \Big(\Big(\frac{\alpha|I||J|}{q^{3/2}}\Big)^{1/r} + 1\Big)\right)\sqrt{q}.$$

If $|I||J| \le q^{3/2}$ we take $\alpha = 1$ and then

$$\begin{aligned} D_{\mathcal{A}}(B) &\le \left(2 + 4\cdot 6^r\Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/r} + 1\Big)\right)\sqrt{q} \\ &\le 6^{r+1}\Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/r} + 1\Big)\sqrt{q} \\ &\le 6^{r+1}\Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/(r+1)} + 1\Big)\sqrt{q}. \end{aligned}$$

If $|I||J| > q^{3/2}$ we take $\alpha = \left(\frac{q^{3/2}}{|I||J|}\right)^{1/(r+1)}$ and then

$$\begin{aligned} D_{\mathcal{A}}(B) &\le \left(2\left(\frac{|I||J|}{q^{3/2}}\right)^{1/(r+1)} + 4\cdot 6^r\Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/(r+1)} + 1\Big)\right)\sqrt{q} \\ &\le 6^{r+1}\Big(\Big(\frac{|I||J|}{q^{3/2}}\Big)^{1/(r+1)} + 1\Big)\sqrt{q}. \end{aligned}$$

□

It should be mentioned that, for the particular case $|I| = |J|$ and q a prime p, Garaev obtained the error term $O(|I|^{2/3}\log^{2/3}(|I|p^{-3/4}+2)+p^{1/2})$. We note that the error term in Theorem 5.1 is smaller than Garaev's error term. Furthermore, in the range $q^{3/2} \ll |I||J| \ll q^{3/2}e^{c(\log\log q)^2}$ our error term is also smaller than the error term $O(q^{1/2}\log^2 q)$ established in [10]. For arbitrary intervals, Theorem 5.1 gives $O(q^{1/2}e^{c\sqrt{\log q}})$, which is only slightly weaker than $O(q^{1/2}\log^2 q)$.

Finally, we remark that the analogy of Theorem 5.1 also holds for some other problems of similar flavor, like estimating $|\{x \in I : \ x^2 \in J\}|$ or $|\{x \in I : \ g^x \in J\}|$. These are achieved by employing suitable Sidon sets.

Acknowledgments: Parts of this work have been presented in several places: Conference on Arithmetic Combinatorics (IAS Princeton 2007), Primera Reunión Conjunta Sociedad Matemática Mexicana-Real Sociedad Matemática Española (Oaxaca, México, 2009) and Discrete Analysis Seminar (University of Cambridge, 2010). I would like to thank these institutions for their hospitality. Also I would like to thank I. Shparlinsky and M. Garaev for useful conversations about this work. Finally I am grateful to the referees for their careful reading and comments.

References

[1] P. Csikvári, K. Gyarmati and A. Sárközy, Density and Ramsey type results on algebraic equations with restricted solution sets, *Combinatorica*, to appear.

[2] M. Z. Garaev, The sum-product estimate for large subsets of prime fields. *Proc. Amer. Math. Soc.* 136 (2008) n.8, 2735-2739.

[3] M. Z. Garaev, On the logarithmic factor in error term estimates in certain additive congruence problems, *Acta Arithmetica* 124 (2006) n.1, 33-40.

[4] M. Z. Garaev and Ka-Lam Kueh, Distribution of special sequences modulo a large prime, *International Journal of Mathematics and Mathematical Sciences*, 50 (2003) n.50, 3189-3194.

[5] M. Z. Garaev and C. Shen, On the size of the set $A(A+1)$, *Mathematische Zeitschrift*, 265 (2010) n.1, 125-132.

[6] V. C. García, A note on an additive problem with powers of a primitive root. *Bol. Soc. Mat. Mexicana* (3) 11 (2005) n.1, 1-4.

[7] K. Gyarmati and A. Sárközy, Equations in finite fields with restricted solution sets, II. (Algebraic equations.), *Acta Math. Hungar.* 119 (2008) n.3, 259-280.

[8] Derrick Hart, Liangpan Li, Chun-Yen Shen, Fourier analysis and expanding phenomena in finite fields. *to appear in Proc. Amer. Math. Soc.*

[9] S. V. Konyagin, Bounds of exponential sums over subgroups and Gauss sums, *Proc 4th Intern. Conf. Modern Problems of Number Theory and Its Applications*, Moscow Lomonosov State Univ., Moscow, (2002) 86-114 (in Russian)

[10] Z. Rudnik and A. Zaharescu, The distribution of spacing between small powers of a primitive root, *Israel Journal of Mathematics* 120 (2000), 271-287.

[11] A. Sárközy, On sums and products on residues modulo p. *Acta Arithmetica* 118 (2005), n.4, 403-409.

[12] A. Sárközy, On products and shifted products of residues modulo p. Proceedings of CANT 2005. *Integers* **8**(2), (2008)

[13] I. D. Shkredov, On monochromatic solutions of some nonlinear equations in $\mathbb{Z}/\mathbb{Z}_p$, *Mathematical Notes* 88 (2010), n.3-4, 603-611.

[14] J. Solymosi, Incidences and the spectra of graphs, *Combinatorial Number Theory and Additive Group Theory, Advanced Courses in Mathematics - CRM Barcelona* Birkhäuser Basel (2009), 299-314.

[15] Vinh. Szemeredi-Trotter type theorem and sum-product estimate in finite fields, *to appear in Europ. J. of Comb.*

Instituto de Ciencias Matemáticas (CSIC-UAM-UC3M-UCM) and Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049, Madrid, España

E-mail address: `franciscojavier.cilleruelo@uam.es`