GENERALIZED SIDON SETS

JAVIER CILLERUELO, IMRE Z. RUZSA, AND CARLOS VINUESA

Abstract. We give asymptotic sharp estimates for the cardinality of a set of residue classes with the property that the representation function is bounded by a prescribed number. We then use this to obtain an analogous result for sets of integers, answering an old question of Simon Sidon.

1. Introduction

A *Sidon set* A in a commutative group is a set with the property that the sums $a_1 + a_2$, $a_i \in A$ are all distinct except when they coincide because of commutativity. We consider the case when, instead of that, a bound is imposed on the number of such representations. When this bound is g, these sets are often called $B_2[g]$ sets. This being both clumsy and ambiguous, we will avoid it, and fix our notation and terminology below.

Our main interest is in sets of integers and residue classes, but we formulate our concepts and some results in a more general setting.

Let G be a commutative group.

Definition 1.1. For $A \subset G$, we define the corresponding *representation function* as

$$r(x) = \sharp\{(a_1, a_2) : a_i \in A,\, a_1 + a_2 = x\}.$$

The *restricted representation function* is

$$r'(x) = \sharp\{(a_1, a_2) : a_i \in A,\, a_1 + a_2 = x,\, a_1 \neq a_2\}.$$

Finally, the *unordered representation function* $r^*(x)$ counts the pairs (a_1, a_2) where (a_1, a_2) and (a_2, a_1) are identified. With an ordering given on G (not necessarily in any connection with the group operation) we can write this as

$$r^*(x) = \sharp\{(a_1, a_2) : a_i \in A,\, a_1 + a_2 = x,\, a_1 \leq a_2\}.$$

These functions are not independent; we have always the equality

$$r^*(x) = r(x) - \frac{r'(x)}{2}$$

and the inequalities

$$r'(x) \leq r(x) \leq 2r^*(x).$$

2000 *Mathematics Subject Classification.* 11B50, 11B75, 11B13, 11P70.

During the preparation of this paper, J. C. and C. V. were supported by Grant MTM 2008-03880 of MYCIT.

I. R. is supported by Hungarian National Foundation for Scientific Research (OTKA), Grants No. K 61908, K 72731.

C. V. would like to thank I. R. for his hospitality, as well as that of the Alfréd Rényi Institute of Mathematics, during his stay in Budapest.

We have $r(x) = r'(x)$ except for $x = 2a$ with $a \in A$. If we are in this last case and there are no elements of order 2 in G, then necessarily $r(x) = r'(x) + 1$, and the quantities are more closely connected:

$$r'(x) = 2\left\lfloor \frac{r(x)}{2} \right\rfloor,\ r^*(x) = \left\lceil \frac{r(x)}{2} \right\rceil.$$

This is the case in $\mathbb{Z}$, or in $\mathbb{Z}_q$ for odd values of q. For even q this is not necessarily true, but both for constructions and estimates the difference seems to be negligible, as we shall see. In a group with lots of elements of order 2, like in $\mathbb{Z}_2^m$, the difference is substantial.

Observe that r and r' make sense in a noncommutative group as well, while r^* does not.

Definition 1.2. We say that A is a *g-Sidon set*, if $r(x) \le g$ for all x. It is a *weak g-Sidon set*, if $r'(x) \le g$ for all x. It is an *unordered g-Sidon set*, if $r^*(x) \le g$ for all x.

Note 1.3. When we have a set of integers $C \subseteq [1, m]$, we say that it is a g-Sidon set $(\mathrm{mod}\ m)$ if the residue classes $\{c\ (\mathrm{mod}\ m) : c \in C\}$ form a g-Sidon set in $\mathbb{Z}_m$.

The strongest possible of these concepts is that of an unordered 1-Sidon set, and this is what is generally simply called a Sidon set. A weak 2-Sidon set is sometimes called a weak Sidon set.

These concepts are closely connected. If there are no elements of order 2, then $2k$-Sidon sets and unordered k-Sidon sets coincide, in particular, a Sidon set is the same as a 2-Sidon set. Also, in this case $(2k+1)$-Sidon sets and weak $2k$-Sidon sets coincide. Specially, a 3-Sidon set and a weak 2-Sidon set are the same.

Our aim is to find estimates for the maximal size of a g-Sidon set in a finite group, or in an interval of integers.

1.1. **The origin of the problem: g-Sidon sets in the integers.** In 1932, the analyst S. Sidon asked to a young P. Erdős about the maximal cardinality of a g-Sidon set of integers in $\{1, \dots, n\}$. Sidon was interested in this problem in connection with the study of the L_p norm of Fourier series with frequencies in these sets but Erdős was captivated by the combinatorial and arithmetical flavour of this problem and it was one of his favorite problems; not in vain it has been one of the main topics in Combinatorial Number Theory.

Definition 1.4. For a positive integer n

$$\beta_g(n) = \max |A| : A \subset \{1, \dots, n\},\ A \text{ is a } g\text{-Sidon set}.$$

We define $\beta'_g(n)$ and $\beta^*_g(n)$ analogously.

The behaviour of this quantity is only known for classical Sidon sets and for weak Sidon sets : we have $\beta_2(n) \sim \sqrt{n}$ and $\beta_3(n) \sim \sqrt{n}$.

The reason which makes easier the case $g = 2$ is that 2-Sidon sets have the property that the differences $a - a'$ are all distinct. Erdős an Turán [5] used this to prove that $\beta_2(n) \le \sqrt{n} + O(n^{1/4})$ and Lindström [9] refined that to get $\beta_2(n) \le \sqrt{n} + n^{1/4} + 1$. For weak Sidon sets Ruzsa [17] proved that $\beta_3(n) \le \sqrt{n} + 4n^{1/4} + 11$.

For the lower bounds, the classical constructions of Sidon sets of Singer [20], Bose [1] and Ruzsa [17] in some finite groups, $\mathbb{Z}_m$, give $\beta_3(n) \geq \beta_2(n) \geq \sqrt{n}(1+o(1))$. Then, $\lim_{n\to\infty} \frac{\beta_2(n)}{\sqrt{n}} = \lim_{n\to\infty} \frac{\beta_3(n)}{\sqrt{n}} = 1$.

However for $g \geq 4$ it has not even been proved that $\lim_{n\to\infty} \beta_g(n)/\sqrt{n}$ exists.

For this reason we write

$$\overline{\beta}_g = \limsup_{n\to\infty} \beta_g(n)/\sqrt{n} \qquad \text{and} \qquad \underline{\beta}_g = \liminf_{n\to\infty} \beta_g(n)/\sqrt{n}.$$

It is very likely that these limits coincide, but this has only been proved for $g = 2, 3$. A wide literature has been written with bounds for $\overline{\beta}_g$ and $\underline{\beta}_g$ for arbitrary g. The trivial counting argument gives $\overline{\beta}_g \leq \sqrt{2g}$ while the strategy of pasting Sidon sets in $\mathbb{Z}_m$ in the obvious way gives $\underline{\beta}_g \geq \sqrt{g/2}$.

The problem of narrowing this gap has attracted the attention of many mathematicians in the last years.

For example, while for $g = 4$ the trivial upper bound gives $\overline{\beta}_4 \leq \sqrt{8}$, it was proved in [2] that $\overline{\beta}_4 \leq \sqrt{6}$, which was refined to $\overline{\beta}_4 \leq 2.3635...$ in [15] and to $\overline{\beta}_4 \leq 2.3218...$ in [7].

On the other hand, Kolountzakis [8] proved that $\underline{\beta}_4 \geq \sqrt{2}$, which was improved to $\underline{\beta}_4 \geq 3/2$ in [3] and to $\underline{\beta}_4 \geq 4/\sqrt{7} = 1.5118...$ in [7].

We describe below the progress done for large g:

$$\begin{aligned} \tfrac{\overline{\beta}_g}{\sqrt{g}} &\leq \sqrt{2} = 1.4142... \text{ (trivial)} \\ &\leq 1.3180... \text{ (J. Cilleruelo - I. Z. Ruzsa - C. Trujillo, [3])} \\ &\leq 1.3039... \text{ (B. Green, [6])} \\ &\leq 1.3003... \text{ (G. Martin - K. O'Bryant, [12])} \\ &\leq 1.2649... \text{ (G. Yu, [22])} \\ &\leq 1.2588... \text{ (G. Martin - K. O'Bryant, [13])} \end{aligned}$$

$$\begin{aligned} \lim_{g\to\infty} \tfrac{\underline{\beta}_g}{\sqrt{g}} &\geq 1/\sqrt{2} = 0.7071... \text{ (M. Kolountzakis, [8])} \\ &\geq 0.75 \text{ (J. Cilleruelo - I. Z. Ruzsa - C. Trujillo, [3])} \\ &\geq 0.7933... \text{ (G. Martin - K. O'Bryant, [11])} \\ &\geq \sqrt{2/\pi} = 0.7978... \text{ (J. Cilleruelo - C. Vinuesa, [4]).} \end{aligned}$$

Our main result connects this problem with a quantity arising from the analogous continuous problem, first studied by Schinzel and Schmidt [18]. Consider all nonnegative real functions f satisfying $f(x) = 0$ for all $x \notin [0, 1]$, and

$$\int_0^1 f(t)f(x-t)\,dt \leq 1$$

for all x. Define the constant σ by

$$\sigma = \sup \int_0^1 f(x)\,dx \tag{1.1}$$

where the supremum is taken over all functions f satisfying the above restrictions.

Theorem 1.5.

$$\lim_{g\to\infty} \frac{\underline{\beta}_g}{\sqrt{g}} = \lim_{g\to\infty} \frac{\overline{\beta}_g}{\sqrt{g}} = \sigma.$$

In other words, the theorem above says that the maximal cardinality of a g-Sidon set in $\{1,\dots,n\}$ is

$$\beta_g(n) = \sigma\sqrt{gn}(1-\varepsilon(g,n))$$

where $\varepsilon(g,n)\to 0$ when both g and n go to infinity.

Schinzel and Schmidt [18] and Martin and O'Bryant [13] conjectured that $\sigma = 2/\sqrt{\pi} = 1.1283...$, and an extremal function was given by $f(x) = 1/\sqrt{\pi x}$ for $0 < x \le 1$. But recently this has been disproved [14] with an explicit f which gives a greater value. The current state of the art for this constant is

$$1.1509... \le \sigma \le 1.2525...$$

both bounds coming from [14].

The main difficulty in Theorem 1.5 is establishing the lower bound for $\lim \frac{\underline{\beta}_g}{\sqrt{g}}$. Indeed the upper bound $\lim \frac{\overline{\beta}_g}{\sqrt{g}} \le \sigma$ was already proved in [4] using a result of Schinzel and Schmidt from [18]. We include however a complete proof of the theorem.

The usual strategy to construct large g-Sidon sets in the integers is pasting large Sidon sets modulo m in a suitable form. The strategy of pasting g-Sidon sets modulo m had not been tried before since there were no large enough known g-Sidon sets modulo m.

Precisely, the heart of the proof of this theorem is the construction of large g-Sidon sets modulo m.

1.2. **g-Sidon sets in finite groups.**

Definition 1.6. For a finite commutative group G write

$$\alpha_g(G) = \max|A| : A\subset G,\ A \text{ is a } g\text{-Sidon set.}$$

We define $\alpha'_g(G)$ and $\alpha^*_g(G)$ analogously. For the cyclic group $G=\mathbb{Z}_q$, with an abuse of notation, we write $\alpha_g(q) = \alpha_g(\mathbb{Z}_q)$.

An obvious estimate of this quantity is

$$\alpha_g(q) \le \sqrt{gq}.$$

Our aim is to show that for large g for some values of q this is asymptotically the correct value. More exactly, write

$$\alpha_g = \limsup_{q\to\infty} \alpha_g(q)/\sqrt{q}.$$

The case $g=2$ (Sidon sets) is well known, we have $\alpha_2 = 1$. It is also known [17] that $\alpha_3 = 1$. Very little is known about α_g for $g\ge 4$.

For $g = 2k^2$, Martin and O'Bryant [11] generalized the well known constructions of Singer [20], Bose [1] and Ruzsa [17], obtaining $\alpha_g \ge \sqrt{g/2}$ for these values of g.

We are unable to exactly determine α_g for any $g\ge 4$, but we will find its asymptotic behaviour. Our main result sounds as follows.

Theorem 1.7. *We have*

$$\alpha_g = \sqrt{g} + O\left(g^{3/10}\right),$$

in particular,

$$\lim_{g\to\infty} \frac{\alpha_g}{\sqrt{g}} = 1.$$

In Section 2, as a warm-up, we give a slight improvement of the obvious upper estimate.

In Section 3 we construct dense g-Sidon sets in groups $\mathbb{Z}_p^2$. In Section 4 we use this to construct g-Sidon sets modulo q for certain values of q.

Section 5 is devoted to the proof of the upper bound of Theorem 1.5. In Section 6 we prove the lower bound of Theorem 1.5 pasting copies of the large g-Sidon sets in $\mathbb{Z}_q$ which we constructed in Section 4. In these two sections, we connect the discrete and the continuous world, combining some ideas from Schinzel and Schmidt and some probabilistic arguments used in [4].

2. An upper estimate

The representation function $r(x)$ behaves differently at elements of $2 \cdot A = \{2a : a \in A\}$ and the rest; in particular, it can be odd only on this set. Hence we formulate our result in a flexible form that takes this into account.

Theorem 2.1. *Let G be a finite commutative group with $|G| = q$. Let $k \geq 2$ and $l \geq 0$ be integers and $A \subset G$ a set such that the corresponding representation function satisfies*

$$r(x) \leq \begin{cases} k, & \text{if } x \notin 2 \cdot A, \\ k + l, & \text{if } x \in 2 \cdot A. \end{cases}$$

We have

$$|A| < \sqrt{(k-1)q} + 1 + \frac{l}{2} + \frac{l(l+1)}{2(k-1)}. \tag{2.1}$$

Corollary 2.2. *Let G be a finite commutative group with $|G| = q$, and let $A \subset G$ be a g-Sidon set. If g is even, then*

$$|A| \leq \sqrt{(g-1)q} + 1.$$

If g is odd, then

$$|A| \leq \sqrt{(g-2)q} + \frac{3}{2} + \frac{1}{g-2}.$$

Indeed, these are cases $k = g$, $l = 0$ and $k = g - 1$, $l = 1$ of the previous theorem.

Corollary 2.3. *Let $A \subset \mathbb{Z}_q$ be a weak g-Sidon set. If q is even, then*

$$|A| \leq \sqrt{(g-1)q} + 2 + \frac{3}{g-1}.$$

If q is odd, then

$$|A| \leq \sqrt{(g-1)q} + \frac{3}{2} + \frac{1}{g-1}.$$

To deduce this, we put $k = g$ and $l = 2$ if q is even, $l = 1$ if q is odd.

Proof. Write $|A| = m$. We shall estimate the quantity

$$R = \sum r(x)^2$$

in two ways.

First, observe that

$$r(x)^2 - kr(x) = r(x)\,(r(x) - k) \le \begin{cases} 0, & \text{if } x \notin 2 \cdot A, \\ l(k+l), & \text{if } x \in 2 \cdot A, \end{cases}$$

hence

$$R \le k \sum r(x) + l(k+l)\,|2 \cdot A|\,.$$

Since clearly $\sum r(x) = m^2$ and $|2 \cdot A| \le m$, we conclude

$$R \le km^2 + l(k+l)m. \tag{2.2}$$

Write

$$d(x) = \sharp\{(a_1, a_2) : a_i \in A,\, a_1 - a_2 = x\}.$$

Clearly $d(0) = m$. We also have $\sum d(x) = m^2$, and, since the equations $x + y = u + v$ and $x - u = v - y$ are equivalent,

$$\sum d(x)^2 = R.$$

We separate the contribution of $x = 0$ and use the inequality of the arithmetic and quadratic mean to conclude

$$R = m^2 + \sum_{x \ne 0} d(x)^2 \ge m^2 + \frac{1}{q-1}\left(\sum_{x \ne 0} d(x)\right)^2 > m^2 + \frac{m^2(m-1)^2}{q}.$$

A comparison with the upper estimate (2.2) yields

$$\frac{m^2(m-1)^2}{q} < (k-1)m^2 + l(k+l)m.$$

This can be rearranged as

$$(m-1)^2 < (k-1)q + \frac{l(k+l)q}{m}.$$

Now if $m < \sqrt{(k-1)q}$, then we are done; if not, we use the opposite inequality to estimate the second summand and we get

$$(m-1)^2 < (k-1)q + \frac{l(k+l)\sqrt{q}}{\sqrt{k-1}}.$$

We take square root and use the inequality $\sqrt{x+y} \le \sqrt{x} + \frac{y}{2\sqrt{x}}$ to obtain

$$m - 1 < \sqrt{(k-1)q} + \frac{l(k+l)}{2(k-1)}$$

which can be written as (2.1). □

3. Construction in certain groups

In this section we construct large g-Sidon sets in groups $G = \mathbb{Z}_p^2$, for primes p. We shall establish the following result.

Theorem 3.1. *Given k, for every sufficiently large prime $p \geq p_0(k)$ there is a set $A \subseteq \mathbb{Z}_p^2$ with $kp - k + 1$ elements which is a g-Sidon set for $g = \lfloor k^2 + 2k^{3/2} \rfloor$.*

Observe that the trivial upper bound in this case is

$$|A| \leq \sqrt{gq} \leq kp\sqrt{1 + \frac{2}{\sqrt{k}}} < (k + \sqrt{k})p.$$

Proof. Let p be a prime. For every $u \not\equiv 0$ in $\mathbb{Z}_p$ consider the set

$$A_u = \left\{ \left(x, \frac{x^2}{u} \right) : x \in \mathbb{Z}_p \right\} \subset \mathbb{Z}_p^2.$$

Clearly $|A_u| = p$.

We are going to study the sumset of two such sets. For any $\underline{a} = (a, b) \in \mathbb{Z}_p^2$ we shall calculate the representation function

$$r_{u,v}(\underline{a}) = \sharp\{(\underline{a}_1, \underline{a}_2) : \underline{a}_1 \in A_u, \underline{a}_2 \in A_v, \underline{a}_1 + \underline{a}_2 = \underline{a}\}.$$

The most important property for us sounds as follows.

Lemma 3.2. *If $u + v \equiv u' + v'$ and $\left(\frac{uvu'v'}{p}\right) = -1$ then $r_{u,v}(x) + r_{u',v'}(x) = 2$ for all x.*

Proof. If $a \equiv x + y$ and $b \equiv \frac{x^2}{u} + \frac{y^2}{v}$, with $uv \not\equiv 0$, then $y \equiv a - x$ and we have $b \equiv \frac{x^2}{u} + \frac{(a-x)^2}{v}$. We can rewrite this equation as $(u + v)x^2 - 2aux + ua^2 - buv \equiv 0$. The discriminant of this quadratic equation is $\Delta \equiv 4uv((u + v)b - a^2)$. The number of solutions is

$$r_{u,v}(a, b) = \begin{cases} 1 & \text{if } \left(\frac{\Delta}{p}\right) = 0 \\ 2 & \text{if } \left(\frac{\Delta}{p}\right) = +1 \quad (\Delta \text{ quadratic residue}) \\ 0 & \text{if } \left(\frac{\Delta}{p}\right) = -1 \quad (\Delta \text{ quadratic nonresidue}). \end{cases}$$

We can express this as

$$r_{u,v}(a, b) = 1 + \left(\frac{\Delta}{p}\right).$$

Now, since

$$\Delta\Delta' \equiv 4uv((u + v)b - a^2)4u'v'((u' + v')b - a^2) \equiv 16uvu'v'((u + v)b - a^2)^2$$

we have

$$\left(\frac{\Delta}{p}\right)\left(\frac{\Delta'}{p}\right) = \left(\frac{\Delta\Delta'}{p}\right) = \left(\frac{uvu'v'}{p}\right)\left(\frac{((u + v)b - a^2)^2}{p}\right) = -\left(\frac{((u + v)b - a^2)^2}{p}\right).$$

If $(u+v)b - a^2 \equiv 0$, we have $\left(\frac{\Delta}{p}\right) = \left(\frac{\Delta'}{p}\right) = 0$. If not, we have $\left(\frac{\Delta}{p}\right)\left(\frac{\Delta'}{p}\right) = -1$. In any case get

$$\left(\frac{\Delta}{p}\right) + \left(\frac{\Delta'}{p}\right) = 0.$$

□

We resume the proof of the theorem.

We put

$$A = \bigcup_{u=t+1}^{t+k} A_u.$$

and we will show that for a suitable choice of t this will be a good set.

Since $(0,0) \in A_u$ for all u and the rest of the A_u's are disjoint, we have $|A| = k(p-1)+1$.

We can estimate the corresponding representation function as

$$r(x) \le \sum_{u,v=t+1}^{t+k} r_{u,v}(x)$$

(equality fails sometimes, because representations involving $(0,0)$ are counted once on the left and several times on the right).

We parametrize the variables of summation as $u = t+i, v = t+j$ with $1 \le i,j \le k$. So $2 \le i+j \le 2k$ and we can write $i+j = k+1+l$ with $|l| \le k-1$.

For fixed l, we have $k-|l|$ pairs i,j (which means $k-|l|$ pairs u,v). These pairs can be split into two groups: n^+ of them will have $\left(\frac{uv}{p}\right) = 1$ and n^- will have $\left(\frac{uv}{p}\right) = -1$. Clearly

$$n^+ + n^- = k - |l|,\ n^+ - n^- = \sum \left(\frac{uv}{p}\right).$$

Of these $n^+ + n^-$ pairs we can combine $\min\{n^+, n^-\}$ into pairs of pairs with opposite quadratic character, that is, with $\left(\frac{uvu'v'}{p}\right) = -1$. For these we use Lemma 3.2 to estimate the sum of the corresponding representation functions $r_{u,v} + r_{u',v'}$ by 2. For the uncoupled pairs we can only estimate the individual values by 2. Altogether this gives

$$\begin{aligned}
\sum_{i+j=k+1+l} r_{u,v}(x) &\le 2(\min\{n^+,n^-\}) + 2(\max\{n^+,n^-\} - \min\{n^+,n^-\}) \\
&= 2(\max\{n^+,n^-\}) \\
&= n^+ + n^- + |n^+ - n^-| \\
&= k - |l| + \left|\sum \left(\frac{uv}{p}\right)\right|.
\end{aligned}$$

Adding this for all possible value of l, for a fixed t we obtain

$$r(x) \le k^2 + \sum_{|l|\le k-1} \left| \sum_{i+j=k+1+l} \left(\frac{(t+i)(t+j)}{p} \right) \right| = k^2 + S_t.$$

We are going to show that S_t is small on average. Since we need values with $u, v \not\equiv 0$, we can use only $0 \le t \le p-1-k$; however, the complete sum is easier to work with. Applying the Cauchy-Schwarz inequality we get

$$\begin{aligned} \sum_{t=0}^{p-1} S_t &= \sum_{t,l} \left| \sum_{i+j=k+1+l} \left(\frac{(t+i)(t+j)}{p} \right) \right| \\ &\le \sqrt{2kp \sum_{l,t} \left(\sum_{i+j=k+1+l} \left(\frac{(t+i)(t+j)}{p} \right) \right)^2} \\ &\le \sqrt{2kp \sum_{i+j=i'+j'} \sum_{t} \left(\frac{(t+i)(t+j)(t+i')(t+j')}{p} \right)}. \end{aligned}$$

To estimate the inner sum we use Weil's Theorem that asserts

$$\left| \sum_{t=0}^{p-1} \left(\frac{f(t)}{p} \right) \right| \le \deg f \sqrt{p}$$

for any polynomial f which is not a constant multiple of a square. Hence

$$\sum_{t=0}^{p-1} \left(\frac{(t+i)(t+j)(t+i')(t+j')}{p} \right) \le 4\sqrt{p}$$

except when the enumerator as a polynomial of t is a square.

The numerator will be a square if the four numbers i, i', j, j' form two equal pairs. This happens exactly $k(2k-1)$ times. Indeed, we may have $i = i'$, $j = j'$, k^2 cases, or $i = j'$, $j = i'$, another k^2 cases. The k cases when all four coincide have been counted twice. Finally, if $i = j$ and $i' = j'$, then the equality of sums implies that all are equal, so this gives no new case. In these cases for the sum we use the trivial upper estimate p.

The total number of quadruples i, i', j, j' is $\le k^3$, since three of them determine the fourth uniquely.

Combining our estimates we obtain

$$\sum_{t=0}^{p-1} S_t \le \sqrt{2p^2k^2(2k-1) + 8p^{3/2}k^4}.$$

This implies that there is a value of t, $0 \le t \le p-k-1$ such that

$$S_t \le \frac{\sqrt{2p^2k^2(2k-1) + 8p^{3/2}k^4}}{p-k} < 2k^{3/2}$$

if p is large enough. This yields that $r(x) < k^2 + 2k^{3/2}$ as claimed. □

4. Construction in certain cyclic groups

In this section we show how to project a set from $\mathbb{Z}_p^2$ into $\mathbb{Z}_q$ with $q = p^2 s$.

Theorem 4.1. *Let $A \subseteq \mathbb{Z}_p^2$ be a g-Sidon set with $|A| = m$, and put $q = p^2 s$ with a positive integer s. There is a g'-Sidon set $A' \subseteq \mathbb{Z}_q$ with $|A'| = ms$ and $g' = g(s+1)$.*

Proof. An element of A is a pair of residues modulo p, which we shall represent by integers in $[0, p-1]$. Given and element $(a, b) \in A$, we put into A' all numbers of the form $a + cp + bsp$ with $0 \le c \le s-1$. Clearly $|A'| = sm$.

To estimate the representation function of A' we need to tell, given a, b, c, how many $a_1, b_1, c_1, a_2, b_2, c_2$ are there such that

$$a + cp + bsp \equiv a_1 + c_1 p + b_1 sp + a_2 + c_2 p + b_2 sp \pmod{p^2 s} \tag{4.1}$$

with $(a_1, b_1), (a_2, b_2) \in A$ and $0 \le c_1, c_2 \le s-1$.

First consider congruence (4.1) modulo p. We have

$$a \equiv a_1 + a_2 \pmod{p},$$

hence $a_1 + a_2 = a + \delta p$ with $\delta = 0$ or 1. We substitute this into (4.1), substract a and divide by p to obtain

$$c + bs \equiv \delta + c_1 + c_2 + (b_1 + b_2)s \pmod{ps}.$$

We take this modulo s:

$$c \equiv \delta + c_1 + c_2 \pmod{s},$$

consequently $\delta + c_1 + c_2 = c + \eta s$ with $\eta = 0$ or 1. Again substituting back, substracting c and dividing by s we obtain

$$b \equiv \eta + b_1 + b_2 \pmod{p}.$$

So $(a, b) = (a_1, b_1) + (a_2, b_2) + (0, \eta)$ which means that for a, b, η given, we have $\le g$ possible values of a_1, b_1, a_2, b_2.

Now we are going to find the number of possible values of c_1, c_2 for $a, b, c, \eta, a_1, b_1, a_2, b_2$ given.

Observe that from these data we can calculate $\delta = (a_1 + a_2 - a)/p$. For c_1, c_2 we have the equation $c_1 + c_2 = c - \delta + \eta s$.

If $\eta = 0$, we have $c_1 \le c$, at most $c + 1$ possibilities.

If $\eta = 1$, we have $c_1 + c_2 \ge c + s - 1$, hence $c - 1 < c_1 \le s - 1$, which gives at most $s - c$ possibilities.

Hence, if a, b, c, η are given, our estimate is $g(c+1)$ or $g(s-c)$, depending on η. Adding the two estimates we get the claimed bound $g(s+1)$. □

On combining this result with Theorem 3.1 we obtain the following result.

Theorem 4.2. *For any positive integers k, s, for every sufficiently large prime p, there is a set $A \subseteq \mathbb{Z}_{p^2 s}$ with $(kp - k + 1)s$ elements which is a $\lfloor k^2 + 2k^{3/2} \rfloor (s+1)$-Sidon set.*

Put $q = p^2 s$ and $g = \lfloor k^2 + 2k^{3/2} \rfloor (s+1)$. Thus,

$$\begin{aligned} \frac{\alpha_g(q)}{\sqrt{gq}} \geq \frac{|A|}{\sqrt{gq}} &= \frac{(kp-k+1)s}{\sqrt{\lfloor k^2+2k^{3/2}\rfloor (s+1)p^2 s}} \\ &\geq \frac{(kp-k)s}{\sqrt{(k^2+2k^{3/2})(s+1)p^2 s}} \\ &\geq \frac{p-1}{p\sqrt{(1+2/\sqrt{k})(1+1/s)}}. \end{aligned}$$

A convenient choice of the parameters is $k = 4s^2$ (so $s = \Theta(g^{1/5})$). Assuming that, we get

$$\frac{\alpha_g(q)}{\sqrt{gq}} \geq \frac{p-1}{p} \cdot \frac{1}{1+1/s}.$$

Thus, the Prime Number Theorem says that

$$\frac{\alpha_g}{\sqrt{g}} = \limsup_{q\to\infty} \frac{\alpha_g(q)}{\sqrt{gq}} \geq \limsup_{p\to\infty} \frac{p-1}{p} \cdot \frac{1}{1+1/s} = 1 + O(g^{-1/5}),$$

which completes the proof of Theorem 1.7.

5. Upper bound

We turn now to the proof of Theorem 1.5, which says:

$$\lim_{g\to\infty} \liminf_{N\to\infty} \frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} = \lim_{g\to\infty} \limsup_{N\to\infty} \frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} = \sigma.$$

We will prove it in two stages:

Part A.

$$\limsup_{g\to\infty} \limsup_{N\to\infty} \frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} \leq \sigma.$$

Part B.

$$\liminf_{g\to\infty} \liminf_{N\to\infty} \frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} \geq \sigma.$$

For Part A we will use the ideas of Schinzel and Schmidt [18], which give a connection between convolutions and number of representations, between the continuous and the discrete world. For the sake of completeness we rewrite the results and the proofs in a more convenient way for our purposes.

Remember from (1.1) the definition of σ:

$$\sigma = \sup_{f\in\mathcal{F}} |f|_1,$$

where $\mathcal{F} = \{f : f \geq 0,\ \mathrm{supp}(f) \subseteq [0,1],\ |f * f|_\infty \leq 1\}$.

We will use the next result, which is assertion (ii) of Theorem 1 in [18] (essentially the same result appears in [13] as Corollary 1.5):

Theorem 5.1. *Let σ be the constant defined above and $\mathcal{Q}_N = \{Q \in \mathbb{R}_{\geq 0}[x] : Q \not\equiv 0, \deg Q < N\}$. Then*

$$\sup_{Q\in\mathcal{Q}_N} \frac{|Q|_1}{\sqrt{N}\sqrt{|Q^2|_\infty}} \leq \sigma,$$

where $|P|_1$ is the sum and $|P|_\infty$ the maximum of the coefficients of a polynomial P.

Proof. First of all, observe that the definition of σ is equivalent to this one:

$$\sigma = \sup_{g\in\mathcal{G}} \frac{|g|_1}{\sqrt{|g*g|_\infty}},$$

where $\mathcal{G} = \{g : g \geq 0,\ \mathrm{supp}(g)\subseteq[0,1]\}$.

Given a polynomial $Q = a_0 + a_1x + \ldots + a_{N-1}x^{N-1}$ in $\mathcal{Q}_N$, we define the step function g with support in $[0,1)$ having

$$g(x) = a_i \ \text{ for } \ \frac{i}{N} \leq x < \frac{i+1}{N} \ \text{ for every } \ i = 0, 1, \ldots, N-1.$$

The convolution of this step function with itself is the polygonal function:

$$g*g(x) = \sum_{i=0}^{j} a_i a_{j-i} \left(x - \frac{j}{N}\right) + \sum_{i=0}^{j-1} a_i a_{j-1-i} \left(\frac{j+1}{N} - x\right) \text{ if } x \in \left[\frac{j}{N}, \frac{j+1}{N}\right)$$

for every $j = 0, 1, \ldots, 2N-1$, where we define $a_N = a_{N+1} = \ldots = a_{2N-1} = 0$.

So,

$$\sup_x (g*g)(x) = \frac{1}{N} \sup_{0\leq j\leq 2N-2} \left(\sum_{i=0}^{j} a_i a_{j-i}\right).$$

Since, obviously, $\int_0^1 g(x)\ dx = \frac{1}{N}\sum_{i=0}^{N-1} a_i$, we have:

$$\frac{|Q|_1}{\sqrt{N}\sqrt{|Q^2|_\infty}} = \frac{\int_0^1 g(x)\ dx}{\sqrt{\sup_x (g*g)(x)}} \leq \sigma.$$

And because we have this for every Q, the theorem is proved. □

Now, given a g-Sidon set $A \subseteq \{0, 1, \ldots, N-1\}$, we define the polynomial $Q_A(x) = \sum_{a\in A} x^a$, so $Q_A^2(x) = \sum_n r(n)x^n$. Then, Theorem 5.1 says that

$$\sigma \geq \frac{|Q_A|_1}{\sqrt{|Q_A^2|_\infty}\sqrt{N}} \geq \frac{|A|}{\sqrt{g}\sqrt{N}}.$$

Since this happens for every g-Sidon set in $\{0, 1, \ldots, N-1\}$, we have that

$$\frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} \leq \sigma.$$

This proves Part A of Theorem 1.5, which is the easy part.

Remark 5.2. In fact, not only Schinzel and Schmidt prove the result above in [18], but they also prove (see Theorem 6.1) that

$$\lim_{N\to\infty} \sup_{Q\in\mathcal{Q}_N} \frac{|Q|_1}{\sqrt{N}\sqrt{|Q^2|_\infty}} = \sigma.$$

Newman polynomials are polynomials all of whose coefficients are 0 or 1. In [22], Gang Yu conjectured that for every sequence of Newman polynomials Q_N with $\deg Q_N = N-1$ and $|Q_N|_1 = o(N)$

$$\limsup_{N\to\infty} \frac{|Q_N|_1}{\sqrt{N}\sqrt{|Q_N^2|_\infty}} \leq 1.$$

Greg Martin and Kevin O'Bryant [13] disproved this conjecture, finding a sequence of Newman polynomials with $\deg Q_N = N-1$, $|Q_N|_1 = o(N)$ and

$$\limsup_{N\to\infty} \frac{|Q_N|_1}{\sqrt{N}\sqrt{|Q_N^2|_\infty}} = \frac{2}{\sqrt{\pi}}.$$

In fact, with the probabilistic method it can be proved without much effort that there is a sequence of Newman polynomials, with $\deg Q_N = N-1$ and $|Q_N|_1 = O(N^{1/2}(\log N)^\beta)$ for any given $\beta > 1/2$, such that

$$\limsup_{N\to\infty} \frac{|Q_N|_1}{\sqrt{N}\sqrt{|Q_N^2|_\infty}} = \sigma.$$

Our Theorem 1.5 says that given $\varepsilon > 0$, there exists a constant c_ε and a sequence of polynomials, Q_N, with $\deg Q_N = N-1$ and $|Q_N|_1 \leq c_\varepsilon N^{1/2}$ such that

$$\limsup_{N\to\infty} \frac{|Q_N|_1}{\sqrt{N}\sqrt{|Q_N^2|_\infty}} \geq \sigma - \varepsilon.$$

Observe that this growth is close to the best possible, since taking $|Q_N|_1 = o(N^{1/2})$ makes $\frac{|Q_N|_1}{\sqrt{N}\sqrt{|Q_N^2|_\infty}} \to 0$.

6. Connecting the discrete and the continuous world

For Part B of the proof of Theorem 1.5 we will need another result of Schinzel and Schmidt (assertion (iii) of Theorem 1 in [18]) which we state in a more convenient form for our purposes:

Theorem 6.1. *For every $0 < \alpha < 1/2$, for any $0 < \varepsilon < 1$ and for every $n > n(\varepsilon)$, there exist non-negative real numbers $a_0, a_1, \ldots, a_n$ such that*

(1) $a_i \leq n^\alpha(1-\varepsilon)$ *for every* $i = 0, 1, \ldots, n$.
(2) $\sum_{i=0}^n a_i \geq n\sigma(1-\varepsilon)$.
(3) $\sum_{0\leq i, m-i\leq n} a_i a_{m-i} \leq n(1+\varepsilon)$ *for every* $m = 0, 1, \ldots, 2n$.

Proof. We start with a real nonnegative function defined in $[0,1]$, g, with $|g * g|_\infty \leq 1$ and $|g|_1$ close to σ, say $|g|_1 \geq \sigma(1-\varepsilon/2)$.

For $r < s$ we have the estimation

$$\begin{aligned}\left(\int_r^s g(x)\ dx\right)^2 &= \int_r^s\int_r^s g(x)g(y)\ dx\ dy \\ &= \int_{r+x}^{s+x}\int_r^s g(x)g(z-x)\ dx\ dz \\ &\le \int_{2r}^{2s}\int_r^s g(x)g(z-x)\ dx\ dz \le 2(s-r)\end{aligned} \tag{6.1}$$

which implies that

$$\int_r^s g(x)\ dx \le \sqrt{2(s-r)}. \tag{6.2}$$

Trying to "discretize" our function g, we define for $i = 0, 1, 2, \ldots, n$:

$$a_i = \frac{n}{2L}\int_{(i-L)/n}^{(i+L)/n} g(x)\ dx$$

where $1 \le L \le n/2$ is an integer that will be determined later.

Estimation (6.2) proves that

$$a_i \le \sqrt{n/L} \ \text{ for } \ i = 0, 1, 2, \ldots, n. \tag{6.3}$$

Now we give a lower bound for the sum $\sum_{i=0}^n a_i$:

$$\sum_{i=0}^n a_i = \frac{n}{2L}\int_0^1 \nu(x)g(x)\ dx,$$

where

$$\begin{aligned}\nu(x) &= \sharp\left\{i \in [0,n] : \frac{i-L}{n} \le x \le \frac{i+L}{n}\right\} \\ &= \sharp\left\{i : \max\{0, nx-L\} \le i \le \min\{n, nx+L\}\right\}.\end{aligned}$$

Taking in account that an interval of length M has $\ge \lfloor M \rfloor$ integers and an interval of length M starting or finishing at an integer has $\lceil M \rceil$ integers, and since $L \in \mathbb{Z}$ and $1 \le L \le n/2$, we have

$$\nu(x) \ge \begin{cases} nx + L = 2L - (L - nx) & \text{if } 0 \le x \le L/n \\ 2L & \text{if } L/n \le x \le 1 - L/n \\ n - nx + L = 2L - (L - n(1-x)) & \text{if } 1 - L/n \le x \le 1 \end{cases}$$

and so

$$\sum_{i=0}^n a_i \ge n\int_0^1 g(x)\ dx - \frac{n}{2L}\int_0^{L/n}(L - nx)g(x)\ dx - \frac{n}{2L}\int_{1-L/n}^1 (L - n(1-x))g(x)\ dx.$$

Now, using the fact that $|g|_1 \ge \sigma(1 - \varepsilon/2)$ and estimation (6.2),

$$\sum_{i=0}^n a_i \ge n\sigma(1 - \varepsilon/2) - \sqrt{2nL}. \tag{6.4}$$

Also, for every $m \leq 2n$ we give an upper bound for the sum $\sum_{0\leq i,m-i\leq n} a_i a_{m-i}$. First we write:

$$\sum_{0\leq i,m-i\leq n} a_i a_{m-i} = \left(\frac{n}{2L}\right)^2 \sum_{0\leq i,m-i\leq n} \int_{(m-i-L)/n}^{(m-i+L)/n} \int_{(i-L)/n}^{(i+L)/n} g(x)g(y)\ dx\ dy.$$

Now, as in (6.1), we set $z = x + y$ and we consider the set:

$$S_i = \left\{(x,z) : \frac{i-L}{n} \leq x \leq \frac{i+L}{n} \ \text{ and } \ \frac{m-i-L}{n} \leq z-x \leq \frac{m-i+L}{n}\right\}.$$

Then,

$$\sum_{0\leq i,m-i\leq n} a_i a_{m-i} = \left(\frac{n}{2L}\right)^2 \sum_{0\leq i,m-i\leq n} \int\int_{S_i} g(x)g(z-x)\ dx\ dz$$

and, defining $\mu(x,z) = \sharp\{\max\{0,m-n\} \leq i \leq \min\{m,n\} : i-L \leq nx \leq i+L \ \text{ and } \ m-i-L \leq n(z-x) \leq m-i+L\}$,

$$\sum_{0\leq i,m-i\leq n} a_i a_{m-i} = \left(\frac{n}{2L}\right)^2 \int\int \mu(x,z) g(x)g(z-x)\ dx\ dz.$$

If we write $h = i - nx$ then we are imposing $-L \leq h \leq L$ and $m - L - nz \leq h \leq m + L - nz$, so

$$-L + \max\{0, m-nz\} \leq h \leq L + \min\{0, m-nz\},$$

and $\mu(x,z) \leq \lambda(z)$, which is the number of h's in this interval (it could be empty), and this number is clearly $\leq 2L+1$. Also, for each fixed h, z moves in an interval of length $2L/n$.

This means (remember that $|g * g|_\infty \leq 1$)

$$\begin{aligned}
\sum_{0\leq i,m-i\leq n} a_i a_{m-i} &\leq \left(\frac{n}{2L}\right)^2 \int \lambda(z) \int g(x)g(z-x)\ dx\ dz \\
&\leq \left(\frac{n}{2L}\right)^2 \int \lambda(z)\ dz \\
&\leq \left(\frac{n}{2L}\right)^2 \frac{2L(2L+1)}{n}
\end{aligned}$$

so the sum

$$\sum_{0\leq i,m-i\leq n} a_i a_{m-i} \leq n\left(1 + \frac{1}{2L}\right). \tag{6.5}$$

Finally, looking at (6.3), (6.4) and (6.5), and choosing the integer $L = \lceil n^{1-2\alpha}/(1-\varepsilon)^2 \rceil$ with $0 < \alpha < 1/2$, for sufficiently large n we'll have:

$$a_i \leq n^\alpha(1-\varepsilon) \qquad , \qquad \sum_{i=0}^{n} a_i \geq n\sigma(1-\varepsilon) \qquad \text{and} \qquad \sum_{0\leq i,m-i\leq n} a_i a_{m-i} \leq n(1+\varepsilon).$$

□

Remark 6.2. Now, we will construct random sets. We want to use the numbers obtained in Theorem 6.1 to define probabilities, p_i, and it will be convenient to know the sum of the p_i's. This is the motivation for defining

$$p_i = a_i \cdot \frac{\sigma n^{1-\alpha}}{\sum_{i=0}^{n} a_i} \quad \text{for} \quad i = 0, 1, \dots, n.$$

Now we fix $\alpha = 1/3$, although any $\alpha \in (0, 1/2)$ would work. Then we have $p_i = a_i \cdot \frac{\sigma n^{2/3}}{\sum_{i=0}^{n} a_i}$, so for any $0 < \varepsilon < 1$ and for every $n > n(\varepsilon)$, we have $p_0, p_1, \dots, p_n$ such that:

$$p_i \leq 1 \quad , \quad \sum_{i=0}^{n} p_i = \sigma n^{2/3} \quad \text{and} \quad \sum_{0 \leq i, m-i \leq n} p_i p_{m-i} \leq n^{1/3} \frac{1+\varepsilon}{(1-\varepsilon)^2}.$$

In order to prove that the number of elements and the number of representations in our probabilistic sets are what we expect with high probability, we'll use Chernoff's inequality (see Corollary 1.9 in [21]).

Proposition 6.3. ***(Chernoff's inequality)*** *Let $X = t_1 + \cdots + t_n$ where the t_i are independent Boolean random variables. Then for any $\delta > 0$*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \delta \mathbb{E}(X)) \leq 2e^{-\min(\delta^2/4, \delta/2)\mathbb{E}(X)}. \tag{6.6}$$

Then, we have the next two lemmas which also appear in [4]:

Lemma 6.4. *We consider the probability space of all the subsets $A \subseteq \{0, 1, \dots, n\}$ defined by $\mathbb{P}(i \in A) = p_i$. With the p_i's defined above, given $0 < \varepsilon < 1$, there exists $n_0(\varepsilon)$ such that, for all $n \geq n_0$,*

$$\mathbb{P}(|A| \geq \sigma n^{2/3}(1-\varepsilon)) > 0.9.$$

Proof. Since $|A|$ is a sum of independent Boolean variables and $\mathbb{E}(|A|) = \sum_{i=0}^{n} p_i = \sigma n^{2/3}$, we can apply Proposition 6.3 to deduce that for large enough n

$$\mathbb{P}(|A| < \sigma n^{2/3}(1-\varepsilon)) \leq 2e^{-\sigma n^{2/3}\varepsilon^2/4} < 0.1.$$

□

Lemma 6.5. *We consider the probability space of all the subsets $A \subseteq \{0, 1, \dots, n\}$ defined by $\mathbb{P}(i \in A) = p_i$. Again for the p_i's defined above, given $0 < \varepsilon < 1$, there exists $n_1(\varepsilon)$ such that, for all $n \geq n_1$,*

$$r(m) \leq n^{1/3} \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^3 \text{ for all } m = 0, 1, \dots, 2n$$

with probability > 0.9.

Proof. Since $r(m) = \sum_{0 \leq i, m-i \leq n} \mathbb{I}(i \in A)\mathbb{I}(m - i \in A)$ is a sum of Boolean variables which are not independent, it is convenient to consider

$$r'(m)/2 = \sum_{\substack{0 \leq i, m-i \leq n \\ i < m/2}} \mathbb{I}(i \in A)\mathbb{I}(m - i \in A)$$

leaving in mind that $r(m) = r'(m) + \mathbb{I}(m/2 \in A)$.

From the independence of the indicator functions, and following the notation introduced in Definition 1.1, the expected value of $r'(m)/2$ is

$$\begin{aligned}\mu_m &= \mathbb{E}(r'(m)/2) = \sum_{\substack{0\le i,m-i\le n\\ i<m/2}} \mathbb{E}(\mathbb{I}(i\in A)\mathbb{I}(m-i\in A))\\ &= \sum_{\substack{0\le i,m-i\le n\\ i<m/2}} p_i p_{m-i} \le \frac{n^{1/3}}{2}\cdot\frac{1+\varepsilon}{(1-\varepsilon)^2},\end{aligned}$$

for every $m = 0, 1, \ldots, 2n$, for n large enough.

If $\mu_m = 0$ then $\mathbb{P}(r'(m)/2 > 0) = 0$, so we can consider the next two cases:

- If $\frac{1}{3}\cdot\frac{n^{1/3}(1+\varepsilon)}{2(1-\varepsilon)^2} \le \mu_m$, we can apply Proposition 6.3 (observe that $\varepsilon < 2$ and then $\varepsilon^2/4 \le \varepsilon/2$) to have

$$\begin{aligned}\mathbb{P}\left(r'(m)/2 \ge \frac{n^{1/3}}{2}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2\right) &\le \mathbb{P}(r'(m)/2 \ge \mu_m(1+\varepsilon))\\ &\le 2\exp\left(-\frac{\varepsilon^2\mu_m}{4}\right)\\ &\le 2\exp\left(-\frac{n^{1/3}\varepsilon^2(1+\varepsilon)}{24(1-\varepsilon)^2}\right)\end{aligned}$$

- If $0 < \mu_m < \frac{1}{3}\cdot\frac{n^{1/3}(1+\varepsilon)}{2(1-\varepsilon)^2}$ then we define $\delta = \frac{n^{1/3}}{2\mu_m}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 - 1$ (observe that $\delta \ge 2$ and then $\delta/2 \le \delta^2/4$) and we can apply Proposition 6.3 to have

$$\begin{aligned}\mathbb{P}\left(r'(m)/2 \ge \frac{n^{1/3}}{2}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2\right) &= \mathbb{P}(r'(m)/2 \ge \mu_m(1+\delta))\\ &\le 2\exp\left(-\frac{\delta\mu_m}{2}\right)\\ &= 2\exp\left(-\frac{n^{1/3}}{4}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 + \frac{\mu_m}{2}\right)\\ &\le 2\exp\left(-\frac{n^{1/3}}{4}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 + \frac{n^{1/3}(1+\varepsilon)}{12(1-\varepsilon)^2}\right)\\ &\le 2\exp\left(-\frac{n^{1/3}}{6}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2\right)\end{aligned}$$

Then,

$$\mathbb{P}\left(r'(m)/2 \geq \frac{n^{1/3}}{2}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 \text{ for some } m\right)$$
$$\leq 4n\left(\exp\left(-\frac{n^{1/3}\varepsilon^2(1+\varepsilon)}{24(1-\varepsilon)^2}\right)+\exp\left(-\frac{n^{1/3}}{6}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2\right)\right)$$

which is < 0.1 for n large enough.

Remembering that $r(m) = r'(m) + \mathbb{I}(m/2 \in A)$,

$$\mathbb{P}\left(r(m) \geq n^{1/3}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^2 + \ \mathbb{I}(m/2 \in A) \ \text{ for some } m\right) < 0.1 \text{ for } n \text{ large enough,}$$

and finally

$$\mathbb{P}\left(r(m) \geq n^{1/3}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^3 \text{ for some } m\right) < 0.1 \text{ for } n \text{ large enough.}$$

□

Lemmas 6.4 and 6.5 imply that, given $0 < \varepsilon < 1$, for $n \geq \max\{n_0, n_1\}$, the probability that our random set A satisfies $|A| \geq \sigma n^{2/3}(1-\varepsilon)$ and $r(m) \leq n^{1/3}\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^3$ for every m is greater than 0.8. In particular, for every $n \geq \max\{n_0, n_1\}$ we have a set $A \subseteq \{0, 1, \ldots, n\}$ satisfying these conditions.

7. From residues to integers

In order to prove Part B of Theorem 1.5, we will also need the next lemma, which allows us to "paste" copies of a g_2-Sidon set in a cyclic group with a dilation of a g_1-Sidon set in the integers.

Lemma 7.1. *Let $A = \{0 = a_1 < \ldots < a_k\}$ be a g_1-Sidon set in $\mathbb{Z}$ and let $C \subseteq [1, q]$ be a g_2-Sidon set $\pmod q$. Then $B = \cup_{i=1}^{k}(C + qa_i)$ is a g_1g_2-Sidon set in $[1, q(a_k + 1)]$ with $k|C|$ elements.*

Proof. Suppose we have $g_1g_2 + 1$ representations of an element as the sum of two

$$b_{1,1} + b_{2,1} = b_{1,2} + b_{2,2} = \cdots = b_{1,g_1g_2+1} + b_{2,g_1g_2+1}.$$

Each $b_{i,j} = c_{i,j} + qa_{i,j}$ in a unique way. Now we can look at the equality modulo q to have

$$c_{1,1} + c_{2,1} = c_{1,2} + c_{2,2} = \cdots = c_{1,g_1g_2+1} + c_{2,g_1g_2+1} \pmod q.$$

Since C is a g_2-Sidon set $\pmod q$, by the pigeonhole principle, there are at least $g_1 + 1$ pairs (c_{1,i_1}, c_{2,i_1}), ..., $(c_{1,i_{g_1+1}}, c_{2,i_{g_1+1}})$ such that:

$$c_{1,i_1} = \cdots = c_{1,i_{g_1+1}} \qquad \text{and} \qquad c_{2,i_1} = \cdots = c_{2,i_{g_1+1}}.$$

So the corresponding a_i's satisfy

$$a_{1,i_1} + a_{2,i_1} = \cdots = a_{1,i_{g_1+1}} + a_{2,i_{g_1+1}},$$

and since A is a g_1-Sidon set, there must be an equality

$$a_{1,k} = a_{1,l} \qquad \text{and} \qquad a_{2,k} = a_{2,l}$$

for some $k, l \in \{i_1, \dots, i_{g_1+1}\}$.

Then, for these k and l we have

$$b_{1,k} = b_{1,l} \qquad \text{and} \qquad b_{2,k} = b_{2,l},$$

which completes the proof. □

With all these weapons, we are ready to finish our proof.

Given $0 < \varepsilon < 1$ we have that:

a) For every large enough g we can define $n = n(g)$ as the least integer such that $g = \lfloor n^{1/3} \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^3 \rfloor$, and such an n exists because $n^{1/3} \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^3$ grows more slowly than n. Observe that $n(g) \to \infty$ when $g \to \infty$.

Now, by lemmas 6.4 and 6.5, there is $g_0 = g_0(\varepsilon)$ such that for every $g_1 \geq g_0$ we can consider $n = n(g_1)$ and we have a g_1-Sidon set $A \subseteq \{0, 1, \dots, n\}$ such that

$$\frac{|A|}{\sqrt{g_1}\sqrt{n+1}} \geq \sigma \sqrt{\frac{n}{n+1}} \cdot \frac{(1-\varepsilon)^{5/2}}{(1+\varepsilon)^{3/2}}.$$

b) By Theorem 4.2, there are $g_2 = g_2(\varepsilon)$, $s = s(\varepsilon)$ and a sequence $q_0 = p_r^2 s$, $q_1 = p_{r+1}^2 s$, $q_2 = p_{r+2}^2 s$, … (where p_i is the i-th prime and $r = r(\varepsilon)$) such that for every $i = 0, 1, 2, \dots$ there is a g_2-Sidon set $A_i \subseteq \mathbb{Z}_{q_i}$ with

$$\frac{|A_i|}{\sqrt{g_2 q_i}} \geq 1 - \varepsilon.$$

So, given $0 < \varepsilon < 1$:

1) For every $g \geq g_0(\varepsilon) g_2(\varepsilon)$ there is a $g_1 = g_1(g)$ such that

$$g_1 g_2 \leq g < (g_1 + 1) g_2,$$

and we have $n = n(g_1)$ with $g_1 = \lfloor n^{1/3} \left(\frac{1-\varepsilon}{1+\varepsilon}\right)^3 \rfloor$ and a g_1-Sidon set $A \subseteq \{0, 1, \dots, n\}$ with

$$\frac{|A|}{\sqrt{g_1}\sqrt{n+1}} \geq \sigma \sqrt{\frac{n}{n+1}} \cdot \frac{(1-\varepsilon)^{5/2}}{(1+\varepsilon)^{3/2}}.$$

2) For any $N \geq (n+1) q_0$, there is an $i = i(N)$ such that

$$(n+1) q_i \leq N < (n+1) q_{i+1},$$

and we have a g_2-Sidon set $\pmod{q_i}$, A_i, with

$$\frac{|A_i|}{\sqrt{g_2 q_i}} \geq 1 - \varepsilon.$$

Then, for any g and N large enough, applying Lemma 7.1 we can construct a g_1g_2-Sidon set from A and A_i with $|A||A_i|$ elements in $[1,N]$.

So we have that $\beta_g(N) \geq \beta_{g_1g_2}(N) \geq |A||A_i|$ and then

$$\begin{aligned}\frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} &\geq \frac{\beta_{g_1g_2}(N)}{\sqrt{(g_1+1)g_2}\sqrt{(n+1)q_{i+1}}} \\ &\geq \frac{|A||A_i|}{\sqrt{g_1g_2}\sqrt{(n+1)q_i}}\sqrt{\frac{g_1}{g_1+1}}\sqrt{\frac{q_i}{q_{i+1}}} \\ &\geq \sigma\frac{(1-\varepsilon)^{7/2}}{(1+\varepsilon)^{3/2}}\sqrt{\frac{n}{n+1}}\sqrt{\frac{g_1}{g_1+1}}\sqrt{\frac{p_{r+i}}{p_{r+i+1}}}.\end{aligned}$$

Finally, as a consequence of the Prime Number Theorem, this means that, given $0<\varepsilon<1$, for g and N large enough

$$\frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} \geq \sigma\frac{(1-\varepsilon)^{9/2}}{(1+\varepsilon)^{3/2}}$$

i. e.

$$\liminf_{g\to\infty}\liminf_{N\to\infty}\frac{\beta_g(N)}{\sqrt{g}\sqrt{N}} \geq \sigma.$$

References

[1] R. C. Bose, *An afine analogue of Singer's theorem*. J. Indian Math. Soc. vol 6, 1-15 (1942)

[2] J. Cilleruelo, *An upper bound for $B_2[2]$ sequences*. J. Combin. Theory, Ser. A, vol 89, n°1, 141-144 (2000).

[3] J. Cilleruelo, I. Z. Ruzsa and C. Trujillo, *Upper and lower bounds for finite $B_h[g]$ sequences*. J. Number Theory 97, 26-34 (2002).

[4] J. Cilleruelo, C. Vinuesa, *$B_2[g]$ sets and a conjecture of Schinzel and Schmidt*. Combinatorics, Probability and Computing, vol 17, n°6 (2008).

[5] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*. J. London Math. Soc. 16, 212-215 (1941).

[6] B. Green, *The number of squares and $B_h[g]$ sets*. Acta Arith. 100, 365-390 (2001).

[7] L. Habsieger and A. Plagne, *Ensembles $B_2[2]$: l'étau se resserre*. Integers 2, Paper A2, 20 pp., electronic (2002).

[8] M. Kolountzakis, *The Density of Sets and the Minimum of Dense Cosine Sums*. J. Number Theory 56, 1, 4-11 (1996).

[9] B. Lindström, *An inequality for B_2-sequences*. J. Combinatorial Theory, 6, 211-212 (1969).

[10] B. Lindström, *$B_h[g]$-sequences from B_h-sequences*. Proc. Amer. Math. Soc. 128, 657-659 (2000).

[11] G. Martin, K. O'Bryant, *Constructions of Generalized Sidon Sets*. Journal of Combinatorial Theory, Series A, Volume 113, Issue 4, 591-607 (2006).

[12] G. Martin, K. O'Bryant, *The Symmetric Subset Problem in Continuous Ramsey Theory*. Experiment. Math., Volume 16, no 2, 145-166 (2007).

[13] G. Martin, K. O'Bryant, *The supremum of autoconvolutions, with applications to additive number theory*. arXiv:0807.5121v2.

[14] M. Matolcsi, C. Vinuesa, *Improved bounds on the supremum of autoconvolutions*. arXiv:0907.1379v2.

[15] A. Plagne, *A new upper bound for $B_2[2]$ sets*. J. Combin. Theory, Ser. A, 93, 378-384 (2001).

[16] A. Plagne, *Recent progress on finite $B_h[g]$ sets*, Proceedings of the Thirty-Second Southeastern International Conference on Combinatorics, Graph Theory and Computing (Baton Rouge, LA, 2001), vol. 153, pp. 49-64 (2001).

[17] I. Z. Ruzsa, *Solving a linear equation in a set of integers.* Acta Arithmetica LXV.3 259-282 (1993).
[18] A. Schinzel, W. M. Schmidt, *Comparison of L^1- and $L^\infty-$ norms of squares of polynomials*, Acta Arithmetica, 104, no 3 (2002).
[19] S. Sidon, *Ein Satz Äuber trigonomietrische Polynome und seine Anwendungen in der Theorie der Fourier-Reihen.* Math. Annalen 106, 536-539 (1932).
[20] J. Singer, *A theorem in finite projective geometry and some applications to number theory.* Trans. Amer. Math. Soc. 43 , 377-385 (1938).
[21] T. Tao, V. Vu, *Additive Combinatorics.* Cambridge Studies in Advanced Mathematics, 105 (2006).
[22] G. Yu, *An upper bound for $B_2[g]$ sets.* J. Number Theory 122, no. 1, 211-220 (2007).

Departamento de Matemáticas. Universidad Autónoma de Madrid, 28049 - Madrid, Spain.

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

Alfréd Rényi Institute of Mathematics, Budapest, Pf. 127, H-1364 Hungary.

E-mail address: ruzsa@renyi.hu

Departamento de Matemáticas. Universidad Autónoma de Madrid, 28049 - Madrid, Spain.

E-mail address: c.vinuesa@uam.es