Submitted exclusively to the London Mathematical Society
doi:10.1112/0000/000000

On Sidon sets and asymptotic bases

Javier Cilleruelo

Abstract

Erdős conjectured the existence of an infinite Sidon sequence of positive integers which is an asymptotic basis of order 3. We progress towards this conjecture in several directions. We prove the conjecture for all cyclic groups $\mathbb{Z}_N$ with N large enough. We also show that there is an infinite $B_2[2]$ sequence which is an asymptotic basis of order 3. Finally, we prove that for any $\varepsilon > 0$ there is a Sidon sequence which is an asymptotic basis of order $3+\varepsilon$, that is, any positive sufficiently large integer n can be written as a sum of 4 elements of the sequence, one of them smaller than n^{ε}.

1. *Introduction*

A sequence of positive integers A is a Sidon basis of order h if all the sums $a+a',\ a \le a',\ a,a' \in A$ are distinct (Sidon property) and if every sufficiently large positive integer n can be written as a sum of h elements of A (asymptotic basis of order h). It is not difficult to prove that there are no Sidon basis of order 2. However, Erdős has formulated the following conjecture [**7**, **10**, **11**]:

Conjecture 1.1. *There is a Sidon basis of order* 3.

Even if we are not able to prove this conjecture we give several results relatively close to it. The first one is the modular version of Conjecture 1.1.

Theorem 1.1. *For each large enough N, the cyclic group $\mathbb{Z}_N$ contains a Sidon set which is a basis of order* 3.

We prove Theorem 1.1 in Section 2. An important ingredient in the proof is a result of Granville, Shparlinski and Zaharescu [**14**, Theorem 1] on distributions in the s-dimensional torus of points coming from curves in $\mathbb{F}_p^r$.

The next result is concerned with $B_2[g]$ sequences, a natural generalization of Sidon sequences.

Definition 1. *A sequence A of positive integers is a $B_2[g]$ sequence if each integer n has at most g representations of the form $n = a + a',\ a \le a',\ a, a' \in A$.*

The $B_2[1]$ sequences are just the Sidon sequences.

2000 *Mathematics Subject Classification* 11B13 (primary), 11B30 (secondary).

Erdős [**7**] showed that there exists a $B_2[g]$ sequence of positive integers which is an asymptotic basis of order 3 for some g, and he asked for the minimum possible g (see also [**19**].) Conjecture 1.1 states that the minimum is $g = 1$. We prove that $g \le 2$.

THEOREM 1.2. *There exists a $B_2[2]$ sequence of positive integers which is an asymptotic basis of order 3.*

We next introduce a new generalisation of the notion of basis that appears in the statement of our strongest approximation to Conjecture 1.1.

DEFINITION 2. *For any $\varepsilon > 0$, we say that a sequence A of positive integers is an asymptotic basis of order $h + \varepsilon$ if every sufficiently large positive integer n can be written as a sum of $h + 1$ elements of A, one of them smaller than n^{ε}:*

$$n = a_1 + \cdots + a_{h+1}, \quad a_1, \ldots, a_{h+1} \in A, \quad a_{h+1} \le n^{\varepsilon}.$$

We say that A is a Sidon basis of order $h + \varepsilon$ if in addition it is a Sidon sequence.

THEOREM 1.3. *For any $\varepsilon > 0$ there exists a Sidon basis of order $3 + \varepsilon$. In other words, for any $\varepsilon > 0$ there exists a Sidon sequence A of positive integers such that every large enough positive integer n can be written as*

$$n = a_1 + a_2 + a_3 + a_4, \quad a_1, a_2, a_3, a_4 \in A, \quad a_4 \le n^{\varepsilon}. \tag{1.1}$$

We mention some previous related results. Deshoulliers and Plagne [**5**] have constructed a Sidon basis of order 7 and Kiss [**20**] has proved the existence of a Sidon basis of order 5. At the time when the first version of this work was posted in Arxiv, Kiss, Rozgonyi and Sándor [**21**] proved the existence of a Sidon basis of order to 4. Even if their result is a corollary of Theorem 1.3, it has independent interest. They use a distinct method that we briefly discuss in Section 3.

Theorems 1.2 and 1.3 are proved with the probabilistic method of Erdős and Renyi [**9**]. In the study of sequences satisfying certain additive properties they considered the probabilistic space $\mathcal{S}(\gamma)$ of sequences of positive integers where all the events $x \in A$ are independent and $\mathbb{P}(x \in A) = x^{-\gamma}$, $\gamma > 0$. A formal construction of these probabilistic spaces appears in [**16**].

An easy application of this method shows that, if $\gamma > 3/4$, then almost all sequences in $\mathcal{S}(\gamma)$ are Sidon sequences (after we remove a finite number of elements from the sequence.) On the other hand Erdős and Tetali [**12**] proved that, if $\gamma < 1 - 1/h$, then almost all sequences are asymptotic bases of order h. Therefore, for any γ in the interval $(3/4, 4/5)$, we have that almost all sequences in $\mathcal{S}(\gamma)$ are simultaneously Sidon sequences and asymptotic bases of order 5. This is the argument used in [**20**].

In order to get bases of order $3 + \varepsilon$ we must choose γ close to $2/3$. In this case the sequences in this probabilistic space are far from being Sidon sequences. Typically there will be infinitely many repeated sums of two elements. A way to circumvent this obstacle is to remove the elements involved in such repetitions to obtain a true Sidon sequence. This general idea, known as the alteration method or deletion technique, is standard in the probabilistic method (see e.g. [**2**]) and has been used previously in a similar context [**3**, **4**, **24**].

The main difficulty that appears when applying the alteration method in our problem is that we have to prevent the destruction of all the representations of the form (1.1) for each integer n. Therefore we must prove that, for each n, the number of removed elements involved in the representation of n in the form (1.1) is small.

As far as Theorem 1.2 is concerned, the standard application of the probabilistic method also proves that, if $\gamma > \frac{g+2}{2g+2}$, then with probability 1 a sequence in $\mathcal{S}(\gamma)$ is a $B_2[g]$ sequence (after we remove a finite number of elements.) Therefore, if $\frac{5}{8} < \gamma < \frac{2}{3}$, a random sequence in $\mathcal{S}(\gamma)$ is, with probability 1, simultaneously a $B_2[3]$ sequence and an asymptotic basis of order three. This result appears in [**2**, Section 8.6]. To get a $B_2[2]$ basis of order 3 we must use a more involved argument.

In Section 3 we explain in more detail our strategy and the new ingredients we introduce: the vectorial sunflowers and the modular Sidon bases.

The Sunflower Lemma discovered by Erdős and Rado [**8**] has many applications. In the probabilistic method it has been used to deal with dependent events when each event can be identified with a set. In our proofs it is more convenient to identify each event with a vector and we must use a vectorial version of the Sunflower Lemma. We refer to [**1**] for a recent study of other variants of sunflowers.

The modular Sidon bases are used as tool to simplify the computations of the expected values of some random variables. See Section 3 for a more detailed explanation.

The proofs of Theorems 1.2 and 1.3 are quite similar, except that the last one is technically more involved. They are proved in sections 4 and 5 respectively. The more technical computations of the expected values of the random variables appearing in the proofs are collected in Section 6.

1.1. *General Notation*

Through the paper we will use the following notation:

- $f(n) \gg g(n)$ means that there exists $C > 0$ such that $f(n) > Cg(n)$ for n large enough. We observe that this includes the possibility that $f(n) = 0$ for a finite number of positive integers n.
- $f(n) = o(g(n))$ means that $f(n)/g(n) \to 0$ as $n \to \infty$.
- We write $o_m(1)$ to mean a quantity tending to 0 as $m \to \infty$.

2. *The modular version of the conjecture.*

The statement of our first result about modular Sidon bases is essentially contained in Theorem 1.1. However it includes the extra condition that s_1, s_2, s_3 are pairwise distinct, which is convenient to be used in the proof of Theorem 1.2. Furthermore the proof is short and has an amusing relation with elliptic curves.

THEOREM 2.1. *There exist infinitely many cyclic groups $\mathbb{Z}_N$ containing a Sidon set $S \subset \mathbb{Z}_N$ and such that any element $x \in \mathbb{Z}_N$ can be written in the form*

$$x = s_1 + s_2 + s_3, \quad s_i \in S \tag{2.1}$$

with s_1, s_2, s_3 pairwise distinct.

Proof. Ruzsa [**23**] observed that for all prime p and g a generator of $\mathbb{F}_p^*$, the set

$$S = \{(x, g^x) : \ x = 0, \dots, p-2\}$$

is a Sidon set in $\mathbb{Z}_{p-1} \times \mathbb{Z}_p$. Since $\mathbb{Z}_{p-1} \times \mathbb{Z}_p$ is isomorphic to $\mathbb{Z}_{(p-1)p}$ the set S provides an easy example of a dense Sidon set in a cyclic group. We will prove that S is also a basis of order 3. In other words, that any element $(a, b) \in \mathbb{Z}_{p-1} \times \mathbb{Z}_p$ can be written as

$$(a, b) = (x_1, g^{x_1}) + (x_2, g^{x_2}) + (x_3, g^{x_3}). \tag{2.2}$$

Indeed we will prove that the number of solutions of (2.2) is exactly the number of points $(U, V) \in \mathbb{F}_p^2,\ V \neq 0$ on the elliptic curve $U^2 = 4V^3 + (bV + g^a)^2$ in $\mathbb{F}_p$.

We count, for any $(a, b) \in \mathbb{Z}_{p-1} \times \mathbb{Z}_p$, the number of solutions (x_1, x_2, x_3) of the system

$$x_1 + x_2 + x_3 \equiv a \pmod{p-1} \tag{2.3}$$
$$g^{x_1} + g^{x_2} + g^{x_3} \equiv b \pmod{p}. \tag{2.4}$$

This can be written as

$$g^{x_1} + g^{x_2} + g^{a-x_1-x_2} \equiv b \pmod{p},$$

which is also equivalent to

$$X + Y + \frac{\lambda}{XY} \equiv b \pmod{p} \tag{2.5}$$

with $XY \neq 0$ with the change of variables $g^{x_1} = X,\ g^{x_2} = Y,\ g^a = \lambda$. Now we make an additional change of variables:

$$X = \frac{2V^2}{U - bV - \lambda}, \qquad Y = -\frac{\lambda}{V}.$$

Since $XY \neq 0$ we have to add the condition $V \neq 0,\ U \neq bV + \lambda$. With these restrictions the change of variables is bijective. Applying this change of variables in (2.5) we get

$$\frac{2V^2}{U - bV - \lambda} - \frac{\lambda}{V} - \frac{U - bV - \lambda}{2V} \equiv b \pmod{p},$$

or equivalently,

$$\frac{2V^2}{U - bV - \lambda} \equiv \frac{U + bV + \lambda}{2V} \pmod{p},$$

which can be written as

$$4V^3 + (bV + \lambda)^2 \equiv U^2 \pmod{p}.$$

Each point of this elliptic curve (except the points $(U, V) = (\pm\lambda, 0)$) corresponds to a solution (X, Y) of (2.5). By Hasse's Theorem [**15**] we know that the elliptic curve has $p + O(\sqrt{p})$ points (U, V).

To complete the proof we have to remove the solutions (x_1, x_2, x_3) of (2.3) such that $x_i = x_j$ for some $i \neq j$. Suppose that $x_1 = x_2$. In this case the equation (2.5) leads to $2X + \frac{\lambda}{X^2} \equiv b \pmod{p}$, which is a cubic equation having at most three solutions. Thus, the number of solutions (x_1, x_2, x_3) of (2.3) with some repeated coordinates is at most 9 and the number of representations of (a, b) as a sum of three pairwise distinct elements of S is $p + O(\sqrt{p})$. □

Corollary 2.1 below is a byproduct of the above proof and it will be used in the proof of Theorem 1.3.

Corollary 2.1. *There exist infinitely many cyclic groups $\mathbb{Z}_N$ containing a Sidon set $S \subset \mathbb{Z}_N$ and such that every element $x \in \mathbb{Z}_N$ can be written in the form*

$$x = s_1 + s_2 + s_3 + s_4, \quad s_i \in S,$$

with s_1, s_2, s_3, s_4 pairwise distinct.

Proof. We will see that the set $S \subset \mathbb{Z}_{p-1} \times \mathbb{Z}_p \cong \mathbb{Z}_{(p-1)p}$ described in Theorem 2.1 satisfies the conditions of Corollary 2.1. From the proof of Theorem 2.1 we know that the number of representations of (a, b) as

$$(a, b) = (x_1, g^{x_1}) + (x_2, g^{x_2}) + (x_3, g^{x_3}) + (0, 1), \quad x_i \neq x_j,\ 1 \leq i < j \leq 3 \tag{2.6}$$

is $p+O(\sqrt{p})$. We observe that all these representations of (a,b) satisfy the conditions of Corollary 2.1 except those with $x_i=0$ for some $i=1,2,3$. In these cases the equation (2.5) is a quadratic equation and the number of these special representations is at most 6 for each (a,b). □

2.1. *Proof of Theorem 1.1*

We need the following weaker version of a result of Granville, Shparlinski and Zaharesku (see Theorem 1 in [**14**]):

THEOREM 2.2. *Let $r\ge 2$ and $s\ge 1$ positive integers. Let $\mathcal{C}$ be a curve of degree d in $\mathbb{F}_p^r$ which is absolutely irreducible in $\mathbb{A}^r(\overline{\mathbb{F}}_p)$. Let $h:\mathcal{C}\to\mathbb{A}^s(\overline{\mathbb{F}}_p)$ be a function $h(X)=(h_1(X),\dots,h_s(X))$ where $h_i(X),\ i=1,\dots,s$ are polynomial functions.*

Assume also that there exists $L=L(p)\to\infty$ such that $c_1=\cdots=c_s=0$ whenever $|c_i|\le L,\ i=1,\dots,s$ and $c_1h_1(X)+\cdots+c_sh_s(X)$ is constant along $\mathcal{C}$.

Under these conditions, the set

$$S=\left\{\left(\frac{h_1(X)}{p},\dots,\frac{h_s(X)}{p}\right):\ X\in\mathcal{C}\right\}$$

is well distributed in $\mathbb{T}^s$ when $p\to\infty$.

Proposition 2.1 is a consequence of Theorem 2.2.

PROPOSITION 2.1. *Let $p\equiv 1 \pmod 3$ be a prime. For any integers r_1,r_2, let $\mathcal{C}_{r_1,r_2}$ be the curve in $\mathbb{F}_p^2$ defined by*

$$x_1^2+x_2^2+(x_1+x_2-r_1)^2\equiv r_2 \pmod p. \tag{2.7}$$

The set

$$S_{r_1,r_2}=\left\{\left(\frac{(x_1)_p}{p},\frac{(x_2)_p}{p},\frac{(x_1^2)_p}{p},\frac{(x_2^2)_p}{p}\right):\ (x_1,x_2)\in\mathcal{C}_{r_1,r_2}\right\} \tag{2.8}$$

is well distributed in $[0,1]^4$ when $p\to\infty$. In particular, given $c>0$, any box $B\subset[0,1]^4$ of size $|B|>c$ contains an element of S_{r_1,r_2} if p is large enough.

Proof. The equation (2.7) can be written as

$$3(2x_1+x_2-r_1)^2+(3x_2-r_1)^2\equiv 6r_2-2r_1^2 \pmod p. \tag{2.9}$$

We first consider the case $6r_2-2r_1^2\not\equiv 0 \pmod p$. In this case the curve (2.9) is absolutely irreducible and we will show that the condition of Theorem 2.2 holds with $L=\sqrt{p}/3$. Suppose that there exists constants c_0,c_1,c_2,c_3,c_4 such that

$$c_1x_1+c_2x_2+c_3x_1^2+c_4x_2^2=c_0 \tag{2.10}$$

for all $(x_1,x_2)\in\mathcal{C}_{r_1,r_2}$. From (2.7) we have

$$x_1^2=x_1(r_1-x_2)-x_2^2+r_1x_2+\frac{r_2-r_1^2}{2}\qquad(\text{ in }\mathbb{F}_p).$$

Substituting x_1^2 in (2.10) by this expression we have that

$$c_1x_1+c_2x_2+c_3\left(x_1(r_1-x_2)-x_2^2+r_1x_2+\frac{r_2-r_1^2}{2}\right)+c_4x_2^2=c_0.$$

which is equivalent to the equation

$$(c_1 + c_3(r_1 - x_2))x_1 = (c_3 - c_4)x_2^2 - (c_3r_1 + c_2)x_2 + c_0 + \frac{c_3(r_1^2 - r_2)}{2}.$$

We can write this in a short way as $P(x_2)x_1 = Q(x_2)$ with

$$\begin{aligned} P(x_2) &= -c_3x_2 + c_1 + c_3r_1, \\ Q(x_2) &= (c_3 - c_4)x_2^2 - (c_3r_1 + c_2)x_2 + c_0 + \frac{c_3(r_1^2 - r_2)}{2}. \end{aligned}$$

Multiplying (2.10) by $4c_3$ and completing squares we get

$$(2c_3x_1 + c_1)^2 + 4c_2c_3x_2 + 4c_3c_4x_2^2 = 4c_3c_0 + c_1^2.$$

Multiplying by $P(x_2)^2$ and using that $P(x_2)x_1 = Q(x_2)$ we have

$$(2c_3Q(x_2) + c_1P(x_2))^2 + P(x_2)^2 \left(4c_2c_3x_2 + 4c_3c_4x_2^2 - 4c_3c_0 - c_1^2\right) = 0.$$

This equality must be satisfied for all x_2 corresponding to a point $(x_1, x_2) \in \mathcal{C}_{r_1,r_2}$. Since the left hand side of the above equality is a polynomial in x_2 of degree less than or equal to 4, this is only possible if it is the zero polynomial. It is easy to check that the coefficient of x_2^4 in the polynomial is

$$4c_3^2(c_3 - c_4)^2 + 4c_3^3c_4 = 4c_3^2(c_3^2 + c_4^2 - c_3c_4).$$

If $c_3 \neq 0$ we have that $c_3^2 + c_4^2 - c_3c_4 \equiv 0 \pmod p$. The inequality

$$\left|c_3^2 + c_4^2 - c_3c_4\right| \leq 3L^2 < p,$$

implies that $c_3^2 + c_4^2 - c_3c_4 = 0$ and therefore $c_3 = c_4 = 0$. Thus $c_3 = 0$ in any case.

Since the equation 2.7 is symmetric in x_1 and x_2, we can proceed in the same way to deduce that $c_4 = 0$. Now we have to consider the possibility that $c_1x_1 + c_2x_2 = c_0$ for any $(x_1, x_2) \in \mathcal{C}_{r_1,r_2}$. But this means that all the solutions of the curve $\mathcal{C}_{r_1,r_2}$ lie on that line, which is impossible unless $c_0 = c_1 = c_2 = 0$.

We have proved that the conditions of Theorem 2.2 are satisfied when $6r_2 - 2r_1^2 \not\equiv 0 \pmod p$ and then the sets S_{r_1,r_2} are well distributed in this case.

Assume now that $6r_2 - 2r_1^2 \equiv 0 \pmod p$.

We observe that in this case the curve (2.9) is not absolutely irreducible. Let ω be a solution of $\omega^2 \equiv -3 \pmod p$, which exists because $p \equiv 1 \pmod 3$. It is easy to check that the points (x_1, x_2) of (2.9) are those satisfying either $3x_2 - r_1 = \omega(2x_1 + x_2 - r_1)$ or $3x_2 - r_1 = -\omega(2x_1 + x_2 - r_1)$ and then the curve (2.9) is indeed the union of the two lines:

$$\begin{aligned} \mathcal{C}_{r_1}^+ &= \{(x_1, x_2) : \ 3x_2 - r_1 = +\omega(2x_1 + x_2 - r_1) : \ x_1, x_2 \in \mathbb{F}_p\} \\ \mathcal{C}_{r_1}^- &= \{(x_1, x_2) : \ 3x_2 - r_1 = -\omega(2x_1 + x_2 - r_1) : \ x_1, x_2 \in \mathbb{F}_p\}. \end{aligned}$$

We use again Theorem 2.2 to prove that the set (2.8) is well distributed when (x_1, x_2) belongs to any one of these two lines. This will be shown for the first line, the proof for the second one being analogous.

It is clear that $\mathcal{C}_{r_1}^+$ is absolutely irreducible because it has degree 1. We will prove that in this case $L = (p/4)^{1/3}$ satisfies the condition of Theorem 2.2. By using that $\omega^2 = -3$, the line $\mathcal{C}_{r_1}^+$ can also be written as

$$\mathcal{C}_{r_1}^+ = \{(x_1, x_2) : \ x_1 = -\tfrac{\omega+1}{2}x_2 + \tfrac{\omega+3}{6}r_1 : \ x_1, x_2 \in \mathbb{F}_p\}.$$

Suppose that there exist constants c_0, c_1, c_2, c_3, c_4 such that

$$c_1x_1 + c_2x_2 + c_3x_1^2 + c_4x_2^2 = c_0$$

for all $(x_1, x_2) \in \mathcal{C}^+_{r_1}$. In this case we would have

$$c_1(-\tfrac{\omega+1}{2}x_2 + \tfrac{\omega+3}{6}r_1) + c_2x_2 + c_3(-\tfrac{\omega+1}{2}x_2 + \tfrac{\omega+3}{6}r_1)^2 + c_4x_2^2 = c_0$$

for all $x_2 \in \mathbb{F}_p$, which is not possible if the coefficient of x_2^2 is not 0. So $c_4 = -c_3\left(\frac{\omega+1}{2}\right)^2 = -c_3\frac{\omega^2+2\omega+1}{4} = -c_3\frac{2\omega-2}{4} = c_3\frac{1-\omega}{2}$ and, by using again that $\omega^2 = -3$, we have that

$$c_4^3 = c_3^3\left(\frac{1-\omega}{2}\right)^3 = c_3^3\frac{1-3\omega+3\omega^2-\omega^3}{8} = c_3^3\frac{1-3\omega-9+3\omega}{8} = -c_3^3.$$

From $c_3^3 + c_4^3 = 0$ and $|c_3|, |c_4| < (p/4)^{1/3}$ we obtain that $c_3 = -c_4 = c_3\frac{\omega-1}{2}$. Thus $c_3 = c_4 = 0$.

Now, the relation

$$c_1(-\tfrac{\omega+1}{2}x_2 + \tfrac{\omega+3}{6}r_1) + c_2x_2 = c_0,$$

can not hold for all x_2 if the coefficient of x_2 is not cero. Hence, $c_2 = c_1\frac{\omega+1}{2}$ and it follows that $c_2^3 = c_1^3\left(\frac{\omega+1}{2}\right)^3 = -c_1^3$. From $c_1^3 + c_2^3 = 0$ and $|c_1|, |c_2| \le (p/4)^{1/3}$ we obtain again that $c_1 = -c_2 = -c_1\frac{\omega+1}{2}$. Thus $c_1 = c_2 = 0$.

We have proved that the conditions of Theorem 2.2 are also satisfied when $6r_2 - 2r_1^2 \equiv 0 \pmod p$ for $\mathcal{C}^+_{r_1}$ (and similarly for $\mathcal{C}^-_{r_1}$.) Therefore, the set S_{r_1,r_2} is well distributed in all the cases. □

2.2. *End of the proof of Theorem 1.1*

Erdős and Turán [**13**] have shown that the set

$$A = \{x + (x^2)_p(2p) : \ x = 0, \dots, p-1\},$$

is a Sidon set of integers for any odd prime p.

For a given N, let p be a prime such that $p \equiv 1 \pmod 3$ and $4p^2 < N < 5p^2$. This prime exists if N is large enough. Since $A \subset [0, 2p^2) \subset [0, N/2)$, the set A is a Sidon set in $\mathbb{Z}_N$. We will prove that A is also a basis of order 3 in $\mathbb{Z}_N$.

We observe that for any integer K, the set of integers of the form

$$r_1 + r_2(2p), \quad K \le r_1, r_2 \le \frac{5p-1}{2} + K \tag{2.11}$$

covers an interval of length $5p^2$. This is clear for $K = 0$ and, by translation, for all K. Since $5p^2 > N$, in order to prove that A is a basis of order 3 in $\mathbb{Z}_N$ it is enough to prove that any element of the form (2.11) can be written as a sum of 3 elements of A. We will take $K = \lceil p/4 \rceil$ throughout the proof.

For each (r_1, r_2) we consider the box $B_{r_1,r_2} \subset [0,1]^4$ of all points (y_1, y_2, y_3, y_4) satisfying the following constraints:

$$\left|y_1 - \frac{r_1}{3p}\right|, \ \left|y_2 - \frac{r_1}{3p}\right|, \ \left|y_3 - \frac{r_2}{3p}\right|, \ \left|y_4 - \frac{r_2}{3p}\right| \le \frac{K}{12p}.$$

We must check that $0 < y_i < 1, \ i = 1, \dots, 4$ and that $B_{r_1,r_2} \subset [0,1]^4$. Indeed, since $p \ge 7$, we have

$$y \le \frac{r_i}{3p} + \frac{K}{12p} \le \frac{\frac{5p-1}{2} + K}{3p} + \frac{K}{12p} < \frac{5p-1}{6p} + \frac{5K}{12p} \le \frac{5p-1}{6p} + \frac{5(p+3)}{48p} \le \frac{45p+7}{48p} < 1,$$

and

$$y \ge \frac{r_1}{3p} - \frac{K}{12p} \ge \frac{K}{3p} - \frac{K}{12p} > 0.$$

The size of the box is $|B_{r_1,r_2}| \ge \left(\frac{K}{6p}\right)^4 > 24^{-4}$. Therefore, Proposition 2.1 implies that, for p large enough, there exists an element, say $\left(\frac{x_1}{p}, \frac{x_2}{p}, \frac{(x_1^2)_p}{p}, \frac{(x_2^2)_p}{p}\right)$, with $0 \le x_1, x_2 \le p -$

1, $(x_1, x_2) \in \mathcal{C}_{r_1,r_2}$ satisfying

$$\left|\frac{x_1}{p} - \frac{r_1}{3p}\right|, \ \left|\frac{x_2}{p} - \frac{r_1}{3p}\right|, \ \left|\frac{(x_1^2)_p}{p} - \frac{r_2}{3p}\right|, \ \left|\frac{(x_2^2)_p}{p} - \frac{r_2}{3p}\right| \le \frac{K}{12p}.$$

Since $(x_1, x_2) \in \mathcal{C}_{r_1,r_2}$, there exists an integer $x_3,\ 0 \le x_3 \le p-1$ satisfying

$$\begin{aligned} x_1 + x_2 + x_3 &\equiv r_1 \pmod p \\ x_1^2 + x_2^2 + x_3^2 &\equiv r_2 \pmod p. \end{aligned}$$

Let m be such that $x_1 + x_2 + x_3 = r_1 + mp$. We have

$$\begin{aligned} |m| &\le \left|\frac{x_1}{p} - \frac{r_1}{3p}\right| + \left|\frac{x_2}{p} - \frac{r_1}{3p}\right| + \left|\frac{x_3}{p} - \frac{r_1}{3p}\right| \\ &\le \frac{K}{12p} + \frac{K}{12p} + \max\left(\frac{r_1}{3p}, 1 - \frac{r_1}{3p}\right) \\ &\le \frac{K}{6p} + \max\left(\frac{5p/2 + K}{3p}, 1 - \frac{K}{3p}\right) \\ &\le \max\left(\frac{5p + 3K}{6p}, 1 - \frac{K}{6p}\right) < 1, \end{aligned}$$

since $K = \lceil \frac{p}{4} \rceil$ and $p \ge 7$. This proves that indeed $x_1 + x_2 + x_3 = r_1$. The same argument proves that $(x_1^2)_p + (x_2^2)_p + (x_3^2)_p = r_2$. Thus we have

$$r_1 + r_2(2p) = x_1 + (x_1^2)_p(2p) + x_2 + (x_2^2)_p(2p) + x_3 + (x_3^2)_p(2p),$$

as we wanted to prove.

3. *The probabilistic method with some new tools*

The proofs of Theorems 1.2 and 1.3 are based on the probabilistic method introduced by Erdős and Renyi [**9**] to study sequences satisfying certain arithmetic properties. The book of Alon and Spencer [**2**] is the most complete reference on the probabilistic method and Halberstam and Roth [**16**] is a classic reference for the probabilistic method applied to sequences of integers.

For a given γ, with $0 < \gamma < 1$, Erdős and Renyi introduced the probabilistic space $\mathcal{S}(\gamma)$ of all sequences of positive integers A such that all the events $x \in A$ are independent and $\mathbb{P}(x \in A) = x^{-\gamma}$.

Generally speaking the goal is to prove that a sequence A in $\mathcal{S}(\gamma)$ satisfies certain arithmetic property (or properties) with high probability. To be more precise, we consider certain families Ω_n of sets of positive integers and the random families

$$\Omega_n(A) = \{\omega \in \Omega_n :\ \omega \subset A\}$$

generated by a random sequence A in $S(\gamma)$. Typically we are interested in the random variable

$$X_n(A) = |\Omega_n(A)| = \sum_{\omega \in \Omega_n} I(\omega \subset A).$$

For example if

$$\Omega_n = \{\omega = \{x_1, x_2, x_3\} :\ x_1 + x_2 + x_3 = n\}, \tag{3.1}$$

the random variable $X_n(A)$ counts the number of representation of n as a sum of three elements of a random sequence A in $\mathcal{S}(\gamma)$. In general we are interested in proving that $X_n(A)$ satisfies a certain property P_n. The standard strategy is to prove that

$$\sum_n \mathbb{P}(X_n(A) \text{ does not satisfies } P_n) < \infty \tag{3.2}$$

and then apply the Borel-Cantelli Lemma to deduce that, with probability 1, the random variable $X_n(A)$ satisfies property P_n for all n large enough.

We will modify the probabilistic space $\mathcal{S}(\gamma)$ to force that all the elements of A lie in some residue classes $s \in C \pmod{N}$ for some $C \subset \mathbb{Z}_N$ satisfying suitable conditions. At the end of this section we explain the advantage of this modification. We will write $x \equiv C \pmod{N}$ to mean that $x \equiv s \pmod{N}$ for some $s \in C$.

Also it is technically more convenient to introduce a parameter m to force that the elements of A are greater than a fixed m. This idea was introduced before in [**4**] and allows us to bound (3.2) by a quantity which is $o_m(1)$. At a later step we take m arbitrarily large.

Definition 3. *Let C be a nonempty set of a cyclic group $\mathbb{Z}_N$. For a given γ, $0 < \gamma < 1$ and a given positive integer m, let $\mathcal{S}_m(\gamma, C_N)$ be the probabilistic space of all sequences of positive integers A such that all the events $x \in A$ are independent and such that*

$$\mathbb{P}(x \in A) = \begin{cases} x^{-\gamma} & \text{if} \quad x \equiv C \pmod{N} \text{ and } x > m \\ 0 & \text{otherwise.} \end{cases}$$

Since $X_n(A)$ is a sum of boolean variables we expect that $X_n(A)$ is concentrated around its expected value, $\mu_n = \mathbb{E}(X_n(A))$, with high probability.

When the variables $I(\omega \in A)$ are independent (the sets $\omega \in \Omega_n$ are disjoint), Chernoff's theorem shows that $X_n(A)$ is strongly concentrated around μ_n. However, when the sets in Ω_n are not disjoint, as in the example (3.1), the study of concentration around the mean is more involved.

It is expected, however, that if the dependent events have small correlation we still have enough concentration. Janson's inequality [**17**] serves our purpose for the lower tail:

Theorem 3.1 (Janson's inequality). *Let Ω be a family of sets and let A be a random subset. Let $X(A) = |\{\omega \in \Omega : \ \omega \subset A\}|$ with finite expected value $\mu = \mathbb{E}(X(A))$. Then*

$$\mathbb{P}(X \leq (1-\varepsilon)\mu) \leq exp\left(-\varepsilon^2\mu^2/(2\mu + \Delta(\Omega))\right)$$

where

$$\Delta(\Omega) = \sum_{\substack{\omega, \omega' \in \Omega \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \subset A)$$

and $\omega \sim \omega'$ means that $\omega \cap \omega' \neq \emptyset$ and $\omega \neq \omega'$. In particular, if $\Delta(\Omega) < \mu$ we have that

$$\mathbb{P}(X \leq \mu/2) \leq \exp\left(-\mu/12\right).$$

To deal with the upper tail Erdős and Tetali [**12**] introduced the Sunflowers trick.

3.1. *Sunflowers and vectorial Sunflowers*

A collection of sets $S_1, \dots, S_k$ forms a sunflower if there exists a set S such that $S_i \cap S_j = S$ whenever $i \neq j$. The sets $S_i \setminus S$ are the petals and S is the core of the sunflower. Erdős and Rado [**8**] proved the following interesting lemma.

Lemma 3.1 (Sunflower Lemma). *Let Ω a family of h-sets. If Ω does not contain a sunflower of k petals then $|\Omega| \leq h!(k-1)^h$.*

We will work with a variant of the Sunflower Lemma which deals with vectors instead of sets. The reason is that in our proofs it will be more convenient to work with families Ω of vectors (instead of sets).

Definition 4. *Given a vector $\overline{x} = (x_1, \dots, x_h)$ we define Set($\overline{x}$) $= \{x_1, \dots, x_h\}$. We say that a collection of k distinct vectors $\overline{x}_j,\ j = 1, \dots, k$ forms a disjoint set of k vectors (k-d.s.v. for short) if Set($\overline{x}_j$) $\cap$ Set($\overline{x}_{j'}$) $= \emptyset$ whenever $j \neq j'$.*

Definition 5. *We say that k distinct vectors with h coordinates form a vectorial sunflower (of k petals) if for some $I \subset [h]$ the following two conditions are satisfied:*

- *For all $i \in I$ all the vectors have the same i-th coordinate.*
- *The set of vectors obtained by removing all the i-th coordinates, $i \in I$, forms a k-d.s.v.*

We say that the vectorial sunflower is of type I.

We observe that a vectorial sunflower (of k petals) of type $I = \emptyset$ is a k-d.s.v. The example below forms a vectorial sunflower (of 4 petals) of type $I = \{2, 5\}$.

$$\begin{aligned} \overline{x}_1 &= (\ 7,\ \ \mathbf{7},\ \ 1,\ 13,\ \mathbf{8}) \\ \overline{x}_2 &= (17,\ \ \mathbf{7},\ \ 6,\ \ 6,\ \mathbf{8}) \\ \overline{x}_3 &= (\ 8,\ \ \mathbf{7},\ 18,\ \ 8,\ \mathbf{8}) \\ \overline{x}_4 &= (11,\ \ \mathbf{7},\ \ 4,\ \ 5,\ \mathbf{8}) \end{aligned}$$

We need a vectorial version of Lemma 3.1.

Lemma 3.2 (Vectorial Sunflower Lemma). *Let Ω be a family of vectors of h coordinates. If Ω does not contain a vectorial sunflower of k petals then*

$$|\Omega| \le h!((h^2 - h + 1)(k - 1))^h.$$

Proof. Suppose that $|\Omega| > h!((h^2 - h + 1)(k - 1))^h$. For any $\overline{x} = (x_1, \dots, x_h) \in \Omega$ we consider the set $\mathrm{Set}_h(\overline{x}) = \{hx_1 + 1, hx_2 + 2, \dots, hx_h + h\}$ and the family $\hat{\Omega} = \{\mathrm{Set}_h(\overline{x}) :\ \overline{x} \in \Omega\}$. The sunflower lemma of Erdős-Rao applied to $\hat{\Omega}$ implies that there exists a classical sunflower with $(h^2 - h + 1)(k - 1) + 1$ petals, say $\mathrm{Set}_h(\overline{x}_1), \dots, \mathrm{Set}_h(\overline{x}_{(h^2-h+1)(k-1)+1})$. It is clear that from these sets we can recover the corresponding vectors $\overline{x}_1, \dots, \overline{x}_{(h^2-h+1)(k-1)+1}$ which satisfy the following conditions:

- There exists $I \subset \{1, \dots, h\}$ such that for each $i \in I$ all the $(h^2 - h + 1)(k - 1) + 1$ vectors have the same i-th coordinate.
- For each $i \notin I$, the i-th coordinates of all these vectors are pairwise distinct.

We observe that the conditions above are not enough to make sure that the vectors form a vectorial sunflower. We will prove, however, that the set $\{\overline{x}_1, \dots, \overline{x}_{(h^2-h+1)(k-1)+1}\}$ contains a vectorial sunflower of k petals.

Fix one vector, say $\overline{x}_1$. We know that if $i \notin I$, the i-th coordinate of $\overline{x}_1$ cannot be equal to the i-th coordinate of a distinct vector. However it may be equal to a different i'-th coordinate ($i' \notin I$) of a distinct vector. We observe that for each $i \notin I$ and for each $i' \notin I,\ i' \neq i$ there is at most one such vector. We remove, for each $i \notin I$ and for each $i' \notin I,\ i' \neq i$, such a vector (if it exists). Thus removing at most $h(h - 1)$ vectors we make sure that for all $i \notin I$, the i-th coordinate of x_1 is not equal to any i'-th coordinate ($i' \notin I,\ i' \neq i$) of a distinct vector.

Now we select a second vector and proceed as above. Since the number of original vectors was $(h(h-1)+1)(k-1)+1$ we can select at least k vectors in this way forming a vectorial sunflower of k petals. $\square$

Typically we will deal with families of vectors Ω and with the corresponding random families $\Omega(A) = \{\overline{x} \in \Omega : \ \mathrm{Set}(\overline{x}) \subset A\}$.

Corollary 3.1. *Let Ω_n be a sequence of families of vectors of h coordinates. Suppose that with probability $1 - o_m(1)$ the random families $\Omega_n(A)$ do not contain vectorial sunflowers of K petals for any n. Then, with probability $1 - o_m(1)$ the inequality*

$$|\Omega_n(A)| \le h!((h^2 - h + 1)(K-1))^h$$

holds for all n.

The following proposition will be used several times in the proofs of Theorems 1.2 and 1.3.

Proposition 3.1. *Let $\{\Omega_n\}$ be a sequence of families of vectors and $\{\Omega_n(A)\}$ the corresponding random family where A is a random sequence in $\mathcal{S}_m(\gamma, C_N)$ for some $C \subset \mathbb{Z}_N$. Suppose that there is $\delta > 0$ such that $\mathbb{E}(|\Omega_n(A)|) \ll (n+m)^{-\delta}$. If $K > 1/\delta$ then*

$$\mathbb{P}(\Omega_n(A) \text{ contains a K-d.s.v. for some } n) = o_m(1).$$

Proof.

$$\begin{aligned}
\mathbb{P}\left(\Omega_n(A) \text{ contains a K-d.s.v.}\right) &\le \sum_{\substack{\overline{x}_1,\dots,\overline{x}_K \in \Omega_n \\ \text{form a } K\text{-d.s.v.}}} \mathbb{P}\big(\mathrm{Set}(\overline{x}_1), \dots, \mathrm{Set}(\overline{x}_K) \subset A\big) \\
&= \sum_{\substack{\overline{x}_1,\dots,\overline{x}_K \in \Omega_n \\ \text{form a } K\text{-d.s.v.}}} \mathbb{P}\big(\mathrm{Set}(\overline{x}_1) \subset A\big) \cdots \mathbb{P}\big(\mathrm{Set}(\overline{x}_K) \subset A\big) \\
&\le \frac{1}{K!}\left(\sum_{\overline{x} \in \Omega_n} \mathbb{P}\big(\mathrm{Set}(\overline{x}) \subset A\big)\right)^K \\
&= \frac{\mathbb{E}(|\Omega_n(A)|)^K}{K!} \ll \frac{(n+m)^{-\delta K}}{K!}.
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbb{P}(\Omega_n(A) \text{ contains a K-d.s.v. for some } n) &\ll \sum_n \mathbb{P}\left(\Omega_n(A) \text{ contains a K-d.s.v. }\right) \\
&\ll \sum_n \frac{(n+m)^{-\delta K}}{K!} = o_m(1).
\end{aligned}$$

$\square$

A powerful tool to control the concentration of sum of random boolean dependent variables is the method invented by Kim and Vu [**18**]. Indeed, it was the main ingredient in [**21**]. However, in order to apply this method to our problems we would need to estimate a large amount of expressions (expectations of derivatives) which would duplicate the length of this paper. Instead we have used the Erdős-Tetali method combined with some tricks as the modular trick.

3.2. *The modular trick*

Before we explain the strategy of the proof of Theorem 1.2 let us mention that we will be dealing with sums of the form

$$\sum_{\substack{\overline{x}=(x_1,\dots,x_8)\\ x_1+x_2+x_3=n\\ x_1+x_4=x_5+x_6=x_7+x_8\\ \{x_1,x_4\}\neq\{x_5,x_6\}\neq\{x_7,x_8\}}} \mathbb{P}(x_1,\dots,x_8\in A), \tag{3.3}$$

where A is a random sequence. If the coordinates of a vector $\overline{x}$ are pairwise distinct, then $\mathbb{P}(x_1,\dots,x_8\in A)=\prod_{i=1}^{8}\mathbb{P}(x_i\in A)$ and the computation of (3.3) is straightforward. Unfortunately we have also to consider those vectors with repeated coordinates. There are many patterns to consider and the computation of the sum above would be hard in a standard probabilistic space $\mathcal{S}(\gamma)$. To reduce this unpleasant task we will restrict the sequences A to belong to some residue classes $s\in C, \pmod N$ for some $C\subset\mathbb{Z}_N$ given in Theorem 2.1. This trick will simplify a lot the case analysis of the possible coincidences between the coordinates in the proofs of Lemmas 6.5 and 6.8.

4. $B_2[2]$ *sequences which are asymptotic basis of order* 3

In this section we prove Theorem 1.2.

4.1. *Strategy of the proof*

We start by fixing a cyclic group $\mathbb{Z}_N$ and a set $C\subset\mathbb{Z}_N$ satisfying the conditions of Theorem 2.1. Throughout this section we will consider the probabilistic space $\mathcal{S}_m(7/11,C_N)$. Actually, any value of γ in the interval $(5/8,2/3)$ would work equally fine. We consider the sequence of sets

$$Q_n=\Big\{\omega=\{x_1,x_2,x_3\}:\ x_1+x_2+x_3=n,\ x_i\not\equiv x_j \pmod N,\ i\neq j\Big\}.$$

Given a sequence of positive integers A we define, for each n, the set

$$Q_n(A)=\{\omega\in Q_n:\ \omega\subset A\}.$$

Definition 6 (Lifting process). *The $B_2[2]$-lifting process of a sequence A consists in removing from A those elements $a_1\in A$ such that there exist $a_2,a_3,a_4,a_5,a_6\in A$ with $a_1+a_2=a_3+a_4=a_5+a_6$ and $\{a_1,a_2\}\neq\{a_3,a_4\}\neq\{a_5,a_6\}$.*

We denote by $A_{B_2[2]}$ the surviving elements of A after this process. The sequence $A_{B_2[2]}$ is clearly a $B_2[2]$ sequence.

We define

$$T_n=\{\overline{x}=(x_1,\dots,x_8):\ \overline{x}\text{ satisfies cond}(T_n)\}\quad\text{where}$$

$$\text{cond}(T_n):=\begin{cases}\{x_1,x_2,x_3\}\in Q_n\\ x_1+x_4=x_5+x_6=x_7+x_8, \qquad \{x_1,x_4\}\neq\{x_5,x_6\}\neq\{x_7,x_8\}\\ x_1\equiv x_5\equiv x_7 \pmod N,\ x_4\equiv x_6\equiv x_8 \pmod N.\end{cases}$$

We define also

$$T_n(A)=\{\overline{x}\in T_n:\ \text{Set}(\overline{x})\subset A\}.$$

We will show that $|T_n(A)|$ is an upper bound for the number of representations of n counted in $Q_n(A)$ that are removed in the $B_2[2]$-lifting process of A defined above.

Suppose that $\omega = \{x_1, x_2, x_3\} \in Q_n(A)$ contains an element, say x_1, which is removed in the $B_2[2]$-lifting process. Then there exist $x_4, x_5, x_6, x_7, x_8 \in A$ such that $x_1 + x_4 = x_5 + x_6 = x_7 + x_8$ with $\{x_1, x_4\} \neq \{x_5, x_6\} \neq \{x_7, x_8\}$. On the other hand, since all $x_i \equiv C \pmod N$ and C is a Sidon set in $\mathbb{Z}_N$, interchanging x_5 with x_6 and x_7 with x_8 if needed, we have that $x_1 \equiv x_5 \equiv x_7 \pmod N$ and $x_4 \equiv x_6 \equiv x_8 \pmod N$. Thus, any $\omega \in Q_n(A)$ removed in the $B_2[2]$-lifting process is counted at least once in $T_n(A)$ and we have

$$|Q_n(A_{B_2[2]})| \geq |Q_n(A)| - |T_n(A)|.$$

Since $A_{B_2[2]}$ is a $B_2[2]$ sequence for any sequence A, the proof of Theorem 1.2 reduces to show that there exists a sequence A such that $|Q_n(A)| \gg n^{\delta}$ for some $\delta > 0$ and for n large enough and such that $|T_n(A)| \ll 1$. We do these tasks in Propositions 4.1 and 4.2.

Proposition 4.1. *With probability 1 we have $|Q_n(A)| \gg n^{1/11}$ for n large enough.*

Proof. We apply Janson's inequality to $\Omega = Q_n$ and $X = |Q_n(A)| = \{\omega \in Q_n : \ \omega \subset A\}$ where A is a random sequence in $\mathcal{S}_m(7/11, c_N)$. In Lemma 6.4 we prove that $\mu_n = \mathbb{E}(Q_n(A)) \gg n^{1/11}$ and in Proposition 6.1 we prove that $\Delta(Q_n) \ll n^{-2/11}$ for

$$\Delta(Q_n) = \sum_{\substack{\omega, \omega' \in Q_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \in A).$$

Therefore, for n large enough, we have that $\Delta_n < \mu_n$ and Janson's inequality implies that

$$\mathbb{P}(|Q_n(A)| \leq \mu_n/2) \leq \exp\left(-\mu_n/12\right).$$

Then, for some $c > 0$, we can write

$$\sum_n \mathbb{P}\left(|Q_n(A)| \leq \mu_n/2\right) < \sum_n \exp\left(-cn^{1/11}\right) < \infty$$

and the Borell-Cantelli Lemma implies that $|Q_n(A)| \geq \mu_n/2 \gg n^{1/11}$ with probability 1 for all n large enough. $\square$

In the proof of Proposition 4.2 we use several times Lemma 4.1. We first introduce the following families of vectors, whose expected values are bounded in Lemma 6.3.

$$\begin{aligned} U_{2r} &= \{\overline{x} = (x_1, x_2) : \ x_1 + x_2 = r, \ x_1 \neq x_2\} \\ V_{2r} &= \{\overline{x} = (x_1, x_2) : \ x_1 - x_2 = r\} \\ W_r &= \{\overline{x} = (x_4, x_5, x_6, x_7, x_8) : \ x_5 + x_6 - x_4 = x_7 + x_8 - x_4 = r, \ x_i \neq x_j\}. \end{aligned} \tag{4.1}$$

Lemma 4.1. *Let X_r be any of the three families in* (4.1). *Then*

$$\mathbb{P}(X_r(A) \text{ contains a 12-d.s.v. for some } r) = o_m(1).$$

Proof. Lemma 6.3 implies that $\mathbb{E}(|X_r(A)|) \ll (r+m)^{-2/11}$. The result follows from Proposition 3.1. $\square$

Proposition 4.2. *With probability $1 - o_m(1)$ we have $|T_n(A)| \leq 10^{26}$ for any n.*

Proof. We first show that the statement follows from the next Claim.

Claim. *With probability $1 - o_m(1)$, the family $T_n(A)$ does not contain vectorial sunflowers of 12 petals for any n.*

Assuming the Claim we can apply Corollary 3.1 to the families T_n to deduce that, with probability $1 - o_m(1)$, we have that $|T_n(A)| \le 8!((8^2 - 8 + 1)11)^8 < 10^{26}$ for all n. Hence the Claim implies Proposition 4.2.

We prove the Claim for the distinct possible types $I \subset \{1, \dots, 8\}$ of the vectorial sunflowers in $T_n(A)$. The types we analyze below will cover all the cases, as we will explain later.

1. $I = \emptyset$. By Lemma 6.5 we have $\mathbb{E}(|T_n(A)|) \ll (n+m)^{-1/11}$. Now, Proposition 3.1 implies $\mathbb{P}(T_n(A)$ has a 12-d.s.v. for some $n) = o_m(1)$.

2. $|I \cap \{1, 2, 3\}| = 1$. Assume that $I \cap \{1, 2, 3\} = \{1\}$, the other two cases being similar. If $T_n(A)$ contains a vectorial sunflower (of 12 petals) of type I for some n (denote by l_1 the common first coordinate) then there is a 12-d.s.v. $\overline{x}_j = (x_{2j}, x_{3j}),\ j = 1, \dots, 12$ such that $x_{2j} + x_{3j} = n - l_1$. Thus, for $r = n - l_1$, $U_{2r}(A)$ contains a 12-d.s.v. and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

3. $|I \cap \{1, 4, 5, 6, 7, 8\}| = 1$. Assume that $I \cap \{1, 4, 5, 6, 7, 8\} = \{1\}$, the other cases being similar. If $T_n(A)$ contains a vectorial sunflower (of 12 petals) of type I for some n (denote by l_1 the common first coordinate) then there is a 12-d.s.v. $\overline{x}_j = (x_{4j}, x_{5j}, x_{6j}, x_{7j}, x_{8j}),\ j = 1, \dots 12$ such that $x_{5j} + x_{6j} = x_{7j} + x_{8j} = l_1 + x_{4j}$. Therefore, for $r = l_1$, $W_r(A)$ contains a 12-d.s.v. and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

4. $|I \cap \{1, 4, 5, 6\}| = 2$ or $|I \cap \{1, 4, 7, 8\}| = 2$ or $|I \cap \{5, 6, 7, 8\}| = 2$. Suppose that $|I \cap \{1, 4, 5, 6\}| = 2$. The other cases are similar. We need to distinguish between two essentially distinct cases:
   i) $I \cap \{1, 4, 5, 6\} = \{1, 4\}$. If $T_n(A)$ contains a vectorial sunflower (of 12 petals) of type I (denote by l_1, l_4 the value of the common coordinates) then there is a 12-d.s.v. $\overline{x}_j = (x_{5j}, x_{6j}),\ j = 1, \dots 12$ such that $x_{5j} + x_{6j} = l_1 + l_4$. Thus, for $r = l_1 + l_4$, $U_{2r}(A)$ has a 12-d.s.v. Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

   ii) $I \cap \{1, 4, 5, 6\} = \{1, 5\}$. If $T_n(A)$ contains a vectorial sunflower (of 12 petals) of type I (denote by l_1, l_5 the value of the common coordinates and assume that $l_1 > l_5$) we have that there is an 12-d.s.v. $\overline{x}_j = (x_{4j}, x_{6j}),\ j = 1, \dots 12$ such that $x_{6j} - x_{4j} = l_1 - l_5$. Hence, for $r = l_1 - l_5$, $V_{2r}(A)$ contains a 12-disjoint set and Lemma 4.1 implies the Claim for vectorial sunflowers of this type.

We observe that the sets I considered in the previous analysis cover all the possible cases. The point is that if the subscripts of one of the equations $x_1 + x_2 + x_3 = n,\ \ x_1 + x_4 = x_5 + x_6$, $x_1 + x_4 = x_7 + x_8,\ \ x_5 + x_6 = x_7 + x_8$ are all but one in I, then a vectorial sunflower of that type I cannot exist. For example the cases such that $|I \cap \{1, 2, 3\}| = 2$ are not possible because if two vectors have the same coordinates x_1, x_2, also x_3 must be the same in both vectors. For example, if $\{4, 5, 6\} \subset I$ then these coordinates must be the same in all the vectors of the sunflower, but also x_1 should be the same in all of them. The reader can check that the types not studied above are of this kind. Thus we have proved the Claim. □

5. *Sidon basis of order $3 + \varepsilon$*

In this section we will prove Theorem 1.3. The proof follows the same steps as the proof of Theorem 1.2 but is a little more involved because we have to distinguish an element $x_4 \le n^{\varepsilon}$.

5.1. *Strategy of the proof of Theorem 1.3*

We start by fixing a cyclic group $\mathbb{Z}_N$ and a set $C \subset \mathbb{Z}_N$ satisfying the conditions of Corollary 2.1. Throughout this section we will consider the probabilistic space $\mathcal{S}_m(\gamma, C_N)$ with

$$\gamma = \frac{2}{3} + \frac{\varepsilon}{9 + 9\varepsilon}.$$

Indeed we could take any γ with $\frac{2+3\varepsilon}{3+4\varepsilon} < \gamma < \frac{2+\varepsilon}{3+\varepsilon}$. We consider the families of sets

$$R_n = \Big\{\omega = \{x_1, x_2, x_3, x_4\} : \ \text{satisfying the conditions } \mathrm{cond}(R_n)\Big\} \quad \text{where}$$

$$\mathrm{cond}(R_n) = \begin{cases} x_1 + x_2 + x_3 + x_4 = n, \\ \min(x_1, x_2, x_3, x_4) \le n^{\varepsilon} \\ x_i \not\equiv x_j \pmod N, \ 1 \le i < j \le 4. \end{cases}$$

Given a sequence of positive integers A we define the families:

$$R_n(A) = \Big\{\omega \in R_n : \ \omega \subset A\Big\}.$$

Definition 7 (Sidon lifting process). *The Sidon lifting process of a sequence A consists in removing from A those elements $a \in A$ such that there exist $a', a'', a''' \in A$ with $a + a' = a'' + a'''$, $\{a + a'\} \neq \{a'', a'''\}$.*

We denote by A_{Sidon} the surviving elements of A after this process.

We define

$$B_n(A) = \{\overline{x} = (x_1, \dots, x_7) : \ x_i \in A, \ \overline{x} \text{ satisfies } \mathrm{cond}(B_n)\} \quad \text{where}$$

$$\mathrm{cond}(B_n) := \begin{cases} \{x_1, x_2, x_3, x_4\} \in R_n \\ x_1 + x_5 = x_6 + x_7, \qquad \{x_1, x_5\} \neq \{x_6, x_7\} \\ x_1 \equiv x_6 \pmod N, \ x_5 \equiv x_7 \pmod N). \end{cases}$$

By a similar argument as the one used in the proof of Theorem 1.2 we can see that $|B_n(A)|$ is an upper bound for the number of the representations of n counted in $R_n(A)$ but removed in the Sidon lifting process of A. Thus,

$$|R_n(A_{\mathrm{Sidon}})| \ge |R_n(A)| - |B_n(A)|.$$

Since A_{Sidon} is a Sidon sequence, to prove Theorem 1.3 it is enough to prove that there exists a sequence A such that $|R_n(A)| \gg n^{\delta}$ for some $\delta > 0$ and for n large enough, and such that $|B_n(A)| \ll 1$.

Proposition 5.1. *With probability 1 we have that $|R_n(A)| \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$ for n large enough.*

Proof. We apply Janson's inequality to $\Omega = R_n$ and $X = |R_n(A)| = \{\omega \in R_n : \ \omega \subset A\}$ where A is a random sequence in $\mathcal{S}_m(\gamma, C_N)$. In Proposition 6.7 we prove that $\mu_n = \mathbb{E}(R_n(A)) \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$ and in Proposition 6.2 that $\Delta(R_n) \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}$ where

$$\Delta(R_n) = \sum_{\substack{\omega, \omega' \in R_n \\ \omega \sim \omega'}} \mathbb{P}(\omega, \omega' \in A).$$

Thus for n large enough we have $\Delta(R_n) < \mu_n$ and Janson's inequality implies that

$$\mathbb{P}(|R_n(A)| \le \mu_n/2) \le \exp\left(-\mu_n/12\right).$$

Then, for some $C > 0$, we have

$$\sum_n \mathbb{P}(|R_n(A)| \leq \mu_n/2) < \sum_n \exp\left(-Cn^{\frac{2\varepsilon^2}{9+9\varepsilon}}\right) < \infty$$

and the Borell-Cantelli lemma implies that with probability 1 we have $|R_n(A)| \geq \mu_n/2 \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$ for all n. This proves Proposition 5.1. $\square$

In the proof of Proposition 5.2 we use several times Lemma 5.1. We first introduce the following families of vectors, whose expected values are bounded in Lemma 6.6.

$$\begin{aligned} U_{2r} &= \{\overline{x} = (x_1, x_2) : \ x_1 + x_2 = r, \ x_1 \neq x_2\} \\ U_{3r} &= \{\overline{x} = (x_1, x_2, x_3) : \ x_1 + x_2 + x_3 = r, \ x_i \neq x_j\} \\ V_{2r} &= \{\overline{x} = (x_1, x_2) : \ x_1 - x_2 = r\} \\ V_{3r} &= \{\overline{x} = (x_1, x_2, x_3) : \ x_1 + x_2 - x_3 = r, \ x_i \neq x_j\}. \end{aligned} \tag{5.1}$$

Lemma 5.1. *Let K a positive integer such that $K > 18/\varepsilon^2$. Then for any of the four families X_r in* (5.1)

$$\mathbb{P}(X_r(A) \text{ contains a } K\text{-d.s.v. for some } r) = o_m(1).$$

Proof. Lemma 6.6 implies $\mathbb{E}(|X_r(A)|) \ll (r+m)^{-\varepsilon/6} \ll (r+m)^{-\varepsilon^2/18}$. Then the result follows from Proposition 3.1. $\square$

Proposition 5.2. *With probability $1 - o_m(1)$ we have $|B_n(A)| \ll 1$.*

Proof. We first show that the statement follows from the next Claim.

Claim: *Let K a positive integer such that $K > 18/\varepsilon^2$. Then $B_n(A)$ does not contain vectorial sunflowers of K petals for all n, with probability $1 - o_m(1)$.*

Assuming the Claim we can apply Corollary 3.1 to the families B_n to deduce that $|B_n(A)| \leq 7!((7^2 - 7 + 1)(K - 1))^7$ for all n, with probability $1 - o_m(1)$.

We prove the Claim for the distinct possible type $I \subset \{1, \ldots, 7\}$ of the vectorial sunflowers in $B_n(A)$. The types we analyze below will cover all the cases. It is clear that if $I \cap \{1, 2, 3, 4\} = \{1, 2, 3\}$, then vectorial sunflowers of type I cannot exists because the conditions on B_n implies that also the 4th-coordinate is common for all vectors. The same argument works for any I such that $|I \cap \{1, 2, 3, 4\}| = 3$ or $|I \cap \{1, 5, 6, 7\}| = 3$. Also it is clear that there do not exist vectorial sunflowers of type $I = \{7\}$. Thus we have to consider the types: $I = \emptyset$, $|I \cap \{1, 2, 3, 4\}| = 1$, $|I \cap \{1, 2, 3, 4\}| = 2$, $|I \cap \{1, 5, 6, 7\}| = 1$ and $|I \cap \{1, 5, 6, 7\}| = 2$.

1. $I = \emptyset$. By Lemma 6.8 $\mathbb{E}(|B_n(A)|) \ll (n+m)^{-\frac{\varepsilon^2}{18}}$. Hence, Proposition 3.1 implies $\mathbb{P}(B_n(A) \text{ has a } K\text{-d.s.v. for some } n) = o_m(1)$.

2. $|I \cap \{1, 2, 3, 4\}| = 1$. Suppose that $I \cap \{1, 2, 3, 4\} = \{1\}$. The other three cases are similar. If $B_n(A)$ contains a vectorial sunflower of K petals of this type for some n (denote by l_1 to the common first coordinate) we have that there exists an K-d.s.v. $\overline{x}_j = (x_{2j}, x_{3j}, x_{4j})$, $\mathrm{Set}(\overline{x}_j) \subset A$, $j = 1, \ldots, K$ such that $x_{2j} + x_{3j} + x_{4j} = n - l_1$. Thus, for $r = n - l_1$, $U_{3r}(A)$ contains a K-d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

3. $|I \cap \{1, 2, 3, 4\}| = 2$. Suppose that $I \cap \{1, 2, 3, 4\} = \{1, 2\}$. The other six cases are similar.

If some $B_n(A)$ contains a vectorial sunflower of K petals of this type for some n (denote by l_1, l_2 the common first and second coordinate) we have that there exists an K-d.s.v. $\overline{x}_j = (x_{3j}, x_{4j}),\ \mathrm{Set}(\overline{x}_j) \subset A,\ j = 1, \dots K$ such that $x_{3j} + x_{4j} = n - l_1 - l_2$. Thus, for $r = n - l_1 - l_2$ we have that $U_{2r}(A)$ contains a K-d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

4. $|I \cap \{1, 5, 6, 7\}| = 1$. Suppose that $I \cap \{1, 5, 6, 7\} = \{1\}$. The other four cases are similar. If some $B_n(A)$ contains a vectorial sunflower of K petals of this type (denote by l_1 the common first coordinate) we have that there exists an K-d.s.v. $\overline{x}_j = (x_{5j}, x_{6j}, x_{7j}),\ \mathrm{Set}(\overline{x}_j) \subset A,\ j = 1, \dots K$ such that $x_{6j} + x_{7j} - x_{5j} = l_1$. Thus, for $r = l_1$ we have that $V_{3r}(A)$ contains a K-d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

5. $|I \cap \{1, 5, 6, 7\}| = 2$. We distinguish two essentially distinct cases:
   i) $I \cap \{1, 5, 6, 7\} = \{1, 5\}$. The case $I \cap \{1, 5, 6, 7\} = \{6, 7\}$ is similar.
   If some $B_n(A)$ contains a vectorial sunflower of K petals of this type (let us denote by l_1, l_5 the common first and fifth coordinates respectively) we have that there exists an K-d.s.v. $\overline{x}_j = (x_{6j}, x_{7j}),\ \mathrm{Set}(\overline{x}_j) \subset A,\ j = 1, \dots K$ such that $x_{6j} + x_{7j} = l_1 + l_5$. Thus, for $r = l_1 + l_5$ we have that $U_{2r}(A)$ contains a K-d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

   ii) $I \cap \{1, 5, 6, 7\} = \{1, 6\}$. The case $I \cap \{1, 5, 6, 7\} = \{5, 7\}$ is similar.
   If some $B_n(A)$ contains a vectorial sunflower of K petals of this type (denote by l_1, l_6 the common first and fifth coordinates and assume that $l_1 > l_6$) then there exists a K-d.s.v. $\overline{x}_j = (x_{5j}, x_{7j}),\ \mathrm{Set}(\overline{x}_j) \subset A,\ j = 1, \dots K$ such that $x_{7j} - x_{5j} = l_1 - l_6$. Therefore, for $r = l_1 - l_6$ we have that $V_{2r}(A)$ contains a K-d.s.v. and Lemma 5.1 implies the Claim for vectorial sunflowers of this type.

□

6. *Expected values*

We define the quantities:

$$\sigma_{\alpha,\beta}(n) = \sum_{\substack{x,y \geq 1 \\ x+y=n}} x^{-\alpha} y^{-\beta} = \sum_{1 \leq x < n} x^{-\alpha} (n-x)^{-\beta},$$

$$\tau_{\alpha,\beta}(n) = \sum_{\substack{x,y \geq 1 \\ x-y=n}} x^{-\alpha} y^{-\beta} = \sum_{1 \leq x} x^{-\alpha} (n+x)^{-\beta}$$

and in general

$$\sigma_{\alpha,\beta}(n; m) = \sum_{\substack{x,y > m \\ x+y=n}} x^{-\alpha} y^{-\beta}, \qquad \tau_{\alpha,\beta}(n; m) = \sum_{\substack{x,y > m \\ x-y=n}} x^{-\alpha} y^{-\beta}.$$

The next Lemma will be applied many times later on. We will write $\overset{*}{\ll}$ to mean that we are using Lemma 6.1 in the inequality. All x_i appearing in this section are positive integers.

Lemma 6.1. *For any $\alpha, \beta < 1$ with $\alpha + \beta > 1$ we have*

i) $\sigma_{\alpha,\beta}(n; m) \ll (n+m)^{1-\alpha-\beta}$. iii) $\sigma_{\alpha,\beta}(n) \ll n^{1-\alpha-\beta}$.

ii) $\tau_{\alpha,\beta}(n; m) \ll (n+m)^{1-\alpha-\beta}$. iv) $\tau_{\alpha,\beta}(n) \ll n^{1-\alpha-\beta}$.

Proof. If $n < 2m$, i) holds because $\sigma_{\alpha,\beta}(n;m) = 0$. If $n \geq 2m$ we have

$$\begin{aligned}\sigma_{\alpha,\beta}(n;m) &\leq \sum_{1\leq x\leq n/2} x^{-\alpha}(n-x)^{-\beta} + \sum_{n/2<x<n} x^{-\alpha}(n-x)^{-\beta} \\ &\ll \sum_{1\leq x\leq n/2} x^{-\alpha}n^{-\beta} + \sum_{n/2<x<n} n^{-\alpha}(n-x)^{-\beta} \\ &\ll n^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}.\end{aligned}$$

To prove ii) we distinguish two cases. If $n < m$ we have

$$\tau_{\alpha,\beta}(n;m) \leq \sum_{x>m} x^{-\alpha}(n+x)^{-\beta} \leq \sum_{x>m} x^{-\alpha-\beta} \ll m^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}.$$

If $n \geq m$ we have

$$\begin{aligned}\tau_{\alpha,\beta}(n;m) \leq \sum_{x>m} x^{-\alpha}(n+x)^{-\beta} &= \sum_{m<x<n} x^{-\alpha}(n+x)^{-\beta} + \sum_{x\geq n} x^{-\alpha}(n+x)^{-\beta} \\ &\ll n^{-\beta}\sum_{1\leq x<n} x^{-\alpha} + \sum_{x\geq n} x^{-\alpha-\beta} \ll n^{1-\alpha-\beta} \ll (n+m)^{1-\alpha-\beta}.\end{aligned}$$

The cases iii) and iv) follow from i) and ii) taking $m = 0$. □

Lemma 6.2. *Let a, b be positive integers. Then for any γ, $1/2 < \gamma < 1$,*

$$\sum_{1\leq x} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} \ll (ab)^{1-2\gamma}.$$

Proof. Suppose that $a < b$ and split the sum:

$$\begin{aligned}S &= \sum_{x\leq b} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} + \sum_{x>b} x^{-\gamma}(x+a)^{-\gamma}(x+b)^{1-2\gamma} \\ &\ll b^{1-2\gamma}\sum_{x\leq b} x^{-\gamma}(x+a)^{-\gamma} + \sum_{x>b} x^{1-4\gamma} \overset{*}{\ll} b^{1-2\gamma}a^{1-2\gamma} + b^{2-4\gamma} \ll (ab)^{1-2\gamma}.\end{aligned}$$

□

6.1. *Expected values in $\mathcal{S}_m(7/11, C_N)$*

Lemma 6.3. *We have*

i) $\mathbb{E}(|U_{2r}(A)|) \ll (r+m)^{-3/11}$.
ii) $\mathbb{E}(|V_{2r}(A)|) \ll (r+m)^{-3/11}$.
iii) $\mathbb{E}(|W_r(A)|) \ll (r+m)^{-2/11}$.

Proof.

$$\mathbb{E}(|U_{2r}(A)|) = \sum_{\substack{x,y>m\\x+y=r}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-3/11}.$$

$$\mathbb{E}(|V_{2r}(A)|) = \sum_{\substack{x,y>m\\x-y=m}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-3/11}.$$

$$\begin{aligned}\mathbb{E}(|W_r(A)|) &\le \sum_{\substack{x_4,x_5,x_6,x_7,x_8>m\\x_5+x_6=x_7+x_8=r+x_4}} (x_4x_5x_6x_7x_8)^{-\gamma} \ll \sum_{x_4\ge 1} x_4^{-\gamma}\Big(\sum_{\substack{x,y>m\\x+y=r+x_4}} (xy)^{-\gamma}\Big)^2\\ &\overset{*}{\ll} \sum_{x_4\ge 1} x_4^{-\gamma}(r+m+x_4)^{2-4\gamma} \overset{*}{\ll} (r+m)^{3-5\gamma} \ll (r+m)^{-2/11}.\end{aligned}$$

□

Lemma 6.4. *$\mathbb{E}(|Q_n(A)|) \gg n^{1/11}$ for n large enough.*

Proof. We have $\mathbb{E}(|Q_n(A)|) = \sum_{\{x_1,x_2,x_3\}\in Q_n} \mathbb{P}(x_1,x_2,x_3\in A) \ge n^{-3\gamma}|Q'_n|$, where

$$Q'_n = \Big\{\{x_1,x_2,x_3\}\in Q_n :\ x_i \equiv S \pmod N,\ x_i > m\Big\}.$$

We observe that $S\subset \mathbb{Z}_N$ is such that $n\equiv s_1+s_2+s_3 \pmod N$ for some pairwise distinct s_1,s_2,s_3. We fix s_1,s_2,s_3 and write $x_i = s_i + Ny_i$ and $l = \frac{n-s_1-s_2-s_3}{N}$. Then $|Q'_n| \ge |Q^*_n|$ where

$$|Q^*_n| = \Big|\Big\{\{y_1,y_2,y_3\} :\ y_1+y_2+y_3 = l :\ y_i > m\Big\}\Big| \asymp l^2 \gg n^2,$$

if $l > 10mN$. Thus, $\mathbb{E}(|Q_n(A)|) \ge n^{-3\gamma}|Q^*_n| \gg n^{-3\gamma+2} \gg n^{1/11}$ for n large enough. □

Proposition 6.1. $\Delta(Q_n) \ll n^{-2/11}$.

Proof. If $\omega\sim\omega'$ with $\omega,\omega'\in Q_n$, both sets have exactly one common element, say x_1. Thus

$$\begin{aligned}\Delta(Q_n) = \sum_{\substack{\omega,\omega'\in Q_n\\ \omega\sim\omega'}} \mathbb{P}(\omega,\omega'\subset A) &\ll \sum_{\substack{1\le x_1,x_2,x_3,x'_2,x'_3\\ x_2+x_3=n-x_1\\ x'_2+x'_3=n-x_1}} (x_1x_2x_3x'_2x'_3)^{-\gamma}\\ &\le \sum_{1\le x_1<n} x_1^{-\gamma}\Big(\sum_{\substack{1\le x,y\\ x+y=n-x_1}} (xy)^{-\gamma}\Big)^2 \overset{*}{\ll} \sum_{1\le x_1<n} x_1^{-\gamma}(n-x_1)^{2-4\gamma}\\ &\overset{*}{\ll} n^{3-5\gamma} \ll n^{-2/11}.\end{aligned}$$

□

Lemma 6.5. $\mathbb{E}(|T_n(A)|) \ll (n+m)^{-1/11}$.

Proof. It is clear that $\mathbb{E}(|T_n(A)|) = 0$ if $n < 3m$, so it is enough to prove that $\mathbb{E}(|T_n(A)|) \ll n^{-1/11}$.

We observe that, if $(x_1,\dots,x_8)\in T_n$, then we have some of the following situations:

i) All x_i are pairwise distinct.

ii) $x_7 = x_8$ and $x_1,x_2,x_3,x_4,x_5,x_6,x_7$ are pairwise distinct.

iii) $x_4 \in \{x_2, x_3\}$ and $x_1, x_2, x_3, x_5, x_6, x_7, x_8$ are pairwise distinct.
iv) $x_6 \in \{x_2, x_3\}$ and $x_1, x_2, x_3, x_4, x_5, x_7, x_8$ are pairwise distinct.
v) $x_8 \in \{x_2, x_3\}$ and $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ are pairwise distinct.

In order to simplify the above conditions we observe that iv) and v) are essentially the same one and that x_2 and x_3 play the same role, so iii) can be substituted by $x_4 = x_2$ and iv) and v) by $x_6 = x_2$. Thus we have

$$\begin{aligned}\mathbb{E}(|T_n(A)|) \ll & \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\\ x_1+x_2+x_3=n\\ x_1+x_4=x_5+x_6=x_7+x_8}} (x_1x_2x_3x_4x_5x_6x_7x_8)^{-\gamma} + \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6,x_7\\ x_1+x_2+x_3=n\\ x_1+x_4=x_5+x_6=2x_7}} (x_1x_2x_3x_4x_5x_6x_7)^{-\gamma}\\ & + \sum_{\substack{1\le x_1,x_2,x_3,x_5,x_6,x_7,x_8\\ x_1+x_2+x_3=n\\ x_1+x_2=x_5+x_6=x_7+x_8}} (x_1x_2x_3x_5x_6x_7x_8)^{-\gamma} + \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_7,x_8\\ x_1+x_2+x_3=n\\ x_1+x_4=x_5+x_2=x_7+x_8}} (x_1x_2x_3x_4x_5x_7x_8)^{-\gamma}\\ & = S_1+S_2+S_3+S_4.\end{aligned}$$

$$\begin{aligned}S_1 \ll & \sum_{1\le x_1,x_4<n} (x_1x_4)^{-\gamma} \sum_{\substack{1\le x_2,x_3\\ x_2+x_3=n-x_1}} (x_2x_3)^{-\gamma} \sum_{\substack{1\le x_5,x_6\\ x_5+x_6=x_1+x_4}} (x_5x_6)^{-\gamma} \sum_{\substack{1\le x_7,x_8\\ x_7+x_8=x_1+x_4}} (x_7x_8)^{-\gamma}\\ \overset{*}{\ll} & \sum_{1\le x_1,x_4<n} (x_1x_4)^{-\gamma}(n-x_1)^{1-2\gamma}(x_1+x_4)^{2-4\gamma} \overset{*}{\ll} \sum_{1\le x_1<n} x_1^{3-6\gamma}(n-x_1)^{1-2\gamma} \overset{*}{\ll} n^{5-8\gamma}.\end{aligned}$$

$$\begin{aligned}S_2 \ll & \sum_{1\le x_1,x_4<n} \left(x_1x_4\tfrac{x_1+x_4}{2}\right)^{-\gamma} \sum_{\substack{1\le x_2,x_3\\ x_2+x_3=n-x_1}} (x_2x_3)^{-\gamma} \sum_{\substack{1\le x_5,x_6\\ x_5+x_6=x_1+x_4}} (x_5x_6)^{-\gamma}\\ \overset{*}{\ll} & \sum_{1\le x_1,x_4<n} (x_1x_4(x_1+x_4))^{-\gamma}\,(n-x_1)^{1-2\gamma}(x_1+x_4)^{1-2\gamma}\\ \ll & \sum_{1\le x_1<n} x_1^{-\gamma}(n-x_1)^{1-2\gamma} \sum_{1\le x_4} x_4^{-\gamma}(x_1+x_4)^{1-3\gamma} \overset{*}{\ll} \sum_{1\le x_1<n} x_1^{2-5\gamma}(n-x_1)^{1-2\gamma} \overset{*}{\ll} n^{4-7\gamma}.\end{aligned}$$

$$\begin{aligned}S_3 \ll & \sum_{\substack{1\le x_1,x_2\\ x_1+x_2<n}} (x_1x_2(n-x_1-x_2))^{-\gamma} \sum_{\substack{x_5,x_6>m\\ x_5+x_6=x_1+x_2}} (x_4x_5)^{-\gamma} \sum_{\substack{x_7,x_8>m\\ x_7+x_8=x_1+x_2}} (x_7x_8)^{-\gamma}\\ \ll & \sum_{1\le l<n} \sum_{\substack{1\le x_1,x_2\\ x_1+x_2=l}} (x_1x_2)^{-\gamma}(n-l)^{-\gamma} \sum_{\substack{x_5,x_6>m\\ x_5+x_6=l}} (x_4x_5)^{-\gamma} \sum_{\substack{x_7,x_8>m\\ x_7+x_8=l}} (x_7x_8)^{-\gamma}\\ \overset{*}{\ll} & \sum_{1\le l<n} (n-l)^{-\gamma} l^{3-6\gamma} \overset{*}{\ll} n^{4-7\gamma}.\end{aligned}$$

The estimate of S_4 is more involved. We observe that given x_1, x_3, x_4 the values of x_2 and x_5 are determined by

$$x_2 = n - x_3 - x_1, \qquad x_5 = x_4 + 2x_1 + x_3 - n.$$

$$\begin{aligned}S_4 \ll & \sum_{1\le x_1<n} \sum_{1\le x_3<n-x_1} \sum_{1\le x_4} x_1^{-\gamma}(n-x_3-x_1)^{-\gamma}x_3^{-\gamma}x_4^{-\gamma}(x_4+2x_1+x_3-n)^{-\gamma} \sum_{\substack{1\le x_7,x_8\\ x_7+x_8=x_1+x_4}} (x_7x_8)^{-\gamma}\\ \overset{*}{\ll} & \sum_{1\le x_1<n} \sum_{1\le x_3<n-x_1} x_3^{-\gamma}x_1^{-\gamma}(n-x_3-x_1)^{-\gamma} \sum_{1\le x_4} x_4^{-\gamma}(x_4+2x_1+x_3-n)^{-\gamma}(x_4+x_1)^{1-2\gamma}.\end{aligned}$$

Now we apply Lemma 6.2 to the last sum and later we write $x_3 = n - x_1 - z,\ z \geq 1$ to get

$$\begin{aligned}
S_4 &\ll \sum_{1\leq x_1<n}\ \sum_{1\leq x_3<n-x_1} x_3^{-\gamma} x_1^{1-3\gamma}(n-x_3-x_1)^{-\gamma}(2x_1+x_3-n)^{1-2\gamma}\\
&\ll \sum_{1\leq x_1<n}\ \sum_{1\leq z<n-x_1} (n-x_1-z)^{-\gamma} x_1^{1-3\gamma} z^{-\gamma}(x_1+z)^{1-2\gamma}\\
&\ll \sum_{1\leq x_1<n} x_1^{2-5\gamma} \sum_{1\leq z<n-x_1} (n-x_1-z)^{-\gamma} z^{-\gamma} \overset{*}{\ll} \sum_{1\leq x_1<n} x_1^{2-5\gamma}(n-x_1)^{1-2\gamma} \overset{*}{\ll} n^{4-7\gamma}.
\end{aligned}$$

□

6.2. *Expected values in* $\mathcal{S}_m(\frac{2}{3}+\frac{\varepsilon}{9+9\varepsilon}, C_N)$

LEMMA 6.6. *We have*

i) $\mathbb{E}(|U_{2r}(A)|) \ll (r+m)^{-1/3}$, ii) $\mathbb{E}(|U_{3r}(A)|) \ll (r+m)^{-\varepsilon/6}$.

iii) $\mathbb{E}(|V_{2r}(A)|) \ll (r+m)^{-1/3}$, iv) $\mathbb{E}(|V_{3r}(A)|) \ll (r+m)^{-\varepsilon/6}$.

Proof.

$$\mathbb{E}(|U_{2r}(A)|) = \sum_{\substack{x,y>m\\ x+y=r}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-1/3}.$$

$$\mathbb{E}(|V_{2r}(A)|) = \sum_{\substack{x,y>m\\ x-y=m}} (xy)^{-\gamma} \ll (r+m)^{1-2\gamma} \ll (r+m)^{-1/3}.$$

$$\begin{aligned}
\mathbb{E}(|U_{3r}(A)|) &\leq \sum_{\substack{x,y,z>m\\ x+y+z=r}} (xyz)^{-\gamma} \leq \sum_{z>0} z^{-\gamma} \sum_{\substack{x,y>m\\ x+y=r-z}} (xy)^{-\gamma}\\
&\overset{*}{\ll} \sum_{z} z^{-\gamma}(r-z+m)^{1-2\gamma} \overset{*}{\ll} (r+m)^{2-3\gamma} \ll (r+m)^{-\varepsilon/6}.\\
\mathbb{E}(|V_{3r}(A)|) &\leq \sum_{\substack{x,y,z>m\\ x+y-z=r}} (xyz)^{-\gamma} = \sum_{z\geq 1} z^{-\gamma} \sum_{\substack{x,y>m\\ x+y=r+z}} (xy)^{-\gamma}\\
&\overset{*}{\ll} \sum_{z\geq 1} z^{-\gamma}(r+z+m)^{1-2\gamma} \overset{*}{\ll} (r+m)^{2-3\gamma} \ll (r+m)^{-\varepsilon/6}.
\end{aligned}$$

□

LEMMA 6.7. $\mathbb{E}(R_n(A)) \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$.

Proof. We have $\mathbb{E}(|R_n(A)|) = \sum_{\{x_1,x_2,x_3,x_4\}\in R_n} \mathbb{P}(x_1,x_2,x_3,x_4 \in A) \geq n^{-(3+\varepsilon)\gamma}|R_n'|$, where

$$R_n' = \{\{x_1,x_2,x_3,x_4\} \in R_n :\ x_i \equiv S \pmod N),\ x_i > m\}.$$

We observe that $S \subset \mathbb{Z}_N$ is such that $n \equiv s_1+s_2+s_3+s_4 \pmod N$ for some pairwise distinct s_1,s_2,s_3,s_4. We fix s_1,s_2,s_3,s_4 and write $x_i = s_i + 256y_i$ and $l = \frac{n-s_1-s_2-s_3-s_4}{N}$. Then $|R_n'| \geq |R_n^*|$ where

$$\begin{aligned}
|R_n^*| &= \#\{\{y_1,y_2,y_3,y_4\} :\ y_1+y_2+y_3+y_4 = l,\ m < \min(y_1,y_2,y_3,y_4) \leq n^{\varepsilon}/256\}\\
&\asymp n^{\varepsilon} l^2 \asymp n^{2+\varepsilon}
\end{aligned}$$

Thus, $\mathbb{E}(|R_n(A)|) \geq n^{-(3+\varepsilon)\gamma}|R_n^*| \gg n^{-(3+\varepsilon)\gamma+2+\varepsilon} \gg n^{\frac{2\varepsilon^2}{9+9\varepsilon}}$. □

Proposition 6.2. $\Delta(R_n) \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}$.

Proof. We have that

$$\Delta(R_n) = \sum_{\substack{\omega,\omega' \in R_n \\ \omega \sim \omega'}} \mathbb{P}(\omega,\omega' \in A)$$

We split $\Delta(R_n)$ in several sums according to the common elements of ω and ω'. We suppose that $i = 1$ is the index for which $x_1 \le n^\varepsilon$.

1. $\omega \cap \omega' = \{x_1\}$

$$\sum_{\substack{1\le x_1,x_2,x_3,x_4,x_2',x_3',x_4' \\ x_1+x_2+x_3+x_4=n \\ x_1+x_2'+x_3'+x_4'=n \\ x_1 \le n^\varepsilon}} (x_1x_2x_3x_4x_2'x_3'x_4')^{-\gamma} \ll \sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}\Big(\sum_{\substack{x_2,x_3,x_4 \\ x_2+x_3+x_4=n-x_1}} (x_2x_3x_4)^{-\gamma}\Big)^2$$

$$\overset{*}{\ll} \sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}(n-x_1)^{4-6\gamma} \ll n^{4-6\gamma+\varepsilon(1-\gamma)}.$$

2. $\omega \cap \omega' = \{x_j\}$ for some $j = 2,3,4$. Without lost of generality we consider the case $x_2 = x_2'$:

$$\sum_{\substack{1\le x_1,x_2,x_3,x_4,x_1',x_3',x_4' \\ x_1,x_1' \le n^\varepsilon \\ x_3+x_4=n-x_2-x_1 \\ x_3'+x_4'=n-x_2-x_1'}} (x_1x_2x_3x_4x_1'x_3'x_4')^{-\gamma} \ll \sum_{1\le x_2} x_2^{-\gamma}\Big(\sum_{x_1\le n^\varepsilon} x_1^{-\gamma} \sum_{\substack{1\le x,y \\ x+y=n-x_2-x_1}} (xy)^{-\gamma}\Big)^2$$

Now we split the sum in two sums according to $x_2 \le n - 2n^\varepsilon$ or $n - 2n^\varepsilon < x_2 < n$.

$$\sum_{1\le x_2\le n-2n^\varepsilon} x_2^{-\gamma}\left(\sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}(n-x_2-x_1)^{1-2\gamma}\right)^2 \ll \sum_{1\le x_2\le n} x_2^{-\gamma}\left(\frac{n-x_2}{2}\right)^{2-4\gamma}\left(\sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}\right)^2$$

$$\overset{*}{\ll} n^{3-5\gamma}n^{2(1-\gamma)\varepsilon} \ll n^{(2-2\gamma)\varepsilon+3-5\gamma}.$$

$$\sum_{n-2n^\varepsilon<x_2\le n} x_2^{-\gamma}\left(\sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}(n-x_2-x_1)^{1-2\gamma}\right)^2 \overset{*}{\ll} \sum_{n-2n^\varepsilon<x_2\le n} x_2^{-\gamma}(n-x_2)^{4-6\gamma}$$

$$\ll n^{-\gamma}\sum_{n-2n^\varepsilon<x_2\le n} (n-x_2)^{4-6\gamma} \ll n^{-\gamma+(5-6\gamma)\varepsilon}.$$

3. $\omega \cap \omega' = \{x_1, x_j\}$ for some $j = 2,3,4$. Without lost of generality we consider the case $x_1 = x_1'$ and $x_2 = x_2'$:

$$\sum_{\substack{1\le x_1,x_2,x_3,x_4,x_3',x_4' \\ x_1+x_2+x_3+x_4=n \\ x_1+x_2+x_3'+x_4'=n \\ x_1\le n^\varepsilon}} (x_1x_2x_3x_4x_3'x_4')^{-\gamma} \ll \sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}\sum_{1\le x_2} x_2^{-\gamma}\Big(\sum_{\substack{x_3,x_4 \\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma}\Big)^2$$

$$\overset{*}{\ll} \sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}\sum_{1\le x_2} x_2^{-\gamma}(n-x_1-x_2)^{2-4\gamma}$$

$$\overset{*}{\ll} \sum_{1\le x_1\le n^\varepsilon} x_1^{-\gamma}(n-x_1)^{3-5\gamma} \ll n^{3-5\gamma+\varepsilon(1-\gamma)}.$$

4. $\omega \cap \omega' = \{x_j, x_k\}$ for some $2 \le j < k \le 4$. Without lost of generality we consider the case $x_2 = x_2'$ and $x_3 = x_3'$:

$$\sum_{\substack{1\le x_1,x_2,x_3,x_4,x_1',x_4'\\ x_1+x_2+x_3+x_4=n\\ x_1'+x_2+x_3+x_4'=n\\ x_1,x_1'\le n^{\varepsilon}}} (x_1x_2x_3x_4x_1'x_4')^{-\gamma} \ll \sum_{\substack{1\le x_2,x_3\\ x_2+x_3<n}} (x_2x_3)^{-\gamma}\left(\sum_{1\le x_1\le n^{\varepsilon}} x_1^{-\gamma}(n-x_2-x_3-x_1)^{-\gamma}\right)^2$$

$$\overset{*}{\ll} \sum_{\substack{1\le x_2,x_3\\ x_2+x_3<n-2n^{\varepsilon}}} (x_2x_3)^{-\gamma}\left(n^{-\varepsilon\gamma}\sum_{1\le x_1\le n^{\varepsilon}} x_1^{-\gamma}\right)^2 + \sum_{\substack{1\le x_2,x_3\\ n-2n^{\varepsilon}\le x_2+x_3<n}} (x_2x_3)^{-\gamma}\left((n-x_2-x_3)^{1-2\gamma}\right)^2$$

$$\ll \sum_{\substack{1\le x_2,x_3\\ x_2+x_3<n-2n^{\varepsilon}}} (x_2x_3)^{-\gamma}\left(n^{-\varepsilon\gamma}n^{\varepsilon(1-\gamma)}\right)^2 + \sum_{n-2n^{\varepsilon}\le l<n}\ \sum_{\substack{1\le x_2,x_3\\ x_2+x_3=l}} (x_2x_3)^{-\gamma}(n-l)^{2-4\gamma}$$

$$\overset{*}{\ll} \sum_{\substack{1\le x_2,x_3\\ x_2+x_3<n-2n^{\varepsilon}}} (x_2x_3)^{-\gamma}n^{\varepsilon(2-4\gamma)} + n^{1-2\gamma}\sum_{n-2n^{\varepsilon}\le l<n} (n-l)^{2-4\gamma} \overset{*}{\ll} n^{1-2\gamma+\varepsilon(2-4\gamma)} + n^{1-2\gamma+\varepsilon(3-4\gamma)}.$$

Observe that if $\omega \ne \omega'$ it is not possible that they have three common coordinates. Putting $\gamma = \frac{2}{3} + \frac{\varepsilon}{9+9\varepsilon}$ in each estimate we have that

$$\begin{aligned}\Delta(R_n) &\ll n^{4-6\gamma+\varepsilon(1-\gamma)} + n^{3-5\gamma+\varepsilon(1-\gamma)} + n^{-\gamma+\varepsilon(5-6\gamma)} + n^{1-2\gamma+\varepsilon(2-4\gamma)} + n^{1-2\gamma+\varepsilon(3-4\gamma)}\\ &\ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-5\varepsilon+2\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-6+2\varepsilon+3\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-11\varepsilon-10\varepsilon^2}{9+9\varepsilon}} + n^{\frac{-3-2\varepsilon-\varepsilon^2}{9+9\varepsilon}} \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}}.\end{aligned}$$

□

Lemma 6.8. $\mathbb{E}(B_n(A)) \ll (n+m)^{-\frac{\varepsilon^2}{18}}$.

Proof. It is clear that $\mathbb{E}(|B_n(A)|) = 0$ if $n < 4m$, so it is enough to prove that $\mathbb{E}(B_n(A)) \ll n^{-\frac{\varepsilon^2}{9+9\varepsilon}}$.

We observe that if $(x_1, \dots, x_7) \in B_n(A)$ then some of the following conditions hold:

i) All x_i are pairwise distinct.
ii) $x_6 = x_7$ and all $x_1, x_2, x_3, x_4, x_5, x_6$ are pairwise distinct.
iii) $x_5 \in \{x_2, x_3, x_4\}$ and all $x_1, x_2, x_3, x_4, x_6, x_7$ are pairwise distinct.
iv) $x_6 \in \{x_2, x_3, x_4\}$ and all $x_1, x_2, x_3, x_4, x_5, x_7$ are pairwise distinct.
v) $x_7 \in \{x_2, x_3, x_4,\}$ and all $x_1, x_2, x_3, x_4, x_5, x_6$ are pairwise distinct.

Thus we have

$$\mathbb{E}(|B_n(A)|) \le \sum_{\substack{\overline{x}=(x_1,x_2,x_3,x_4,x_5,x_6,x_7)\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=x_6+x_7\\ \min(x_1,x_2,x_3,x_4)\le n^{\varepsilon}}}{}' \mathbb{P}(x_1, \dots, x_7 \in A)$$

and $\sum'$ means that $\overline{x}$ satisfies i), ii), iii), iv) or v).

In order to simplify these conditions we observe that iv) and v) are essentially the same one, and that x_2, x_3, x_4 play the same role, so iii) can be substituted by $x_5 = x_2$ and iv) and v) by

$x_7 = x_2$. Therefore we have that

$$\begin{aligned}
\mathbb{E}(B_n(A)) \ll & \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6,x_7\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=x_6+x_7\\ \min(x_1,x_2,x_3,x_4)\le n^{\varepsilon}}} (x_1x_2x_3x_4x_5x_6x_7)^{-\gamma} + \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=2x_6\\ \min(x_1,x_2,x_3,x_4)\le n^{\varepsilon}}} (x_1x_2x_3x_4x_5x_6)^{-\gamma}\\
&+ \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_2=x_6+x_7\\ \min(x_1,x_2,x_3,x_4)\le n^{\varepsilon}}} (x_1x_2x_3x_4x_6x_7)^{-\gamma} + \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=x_6+x_2\\ \min(x_1,x_2,x_3,x_4)\le n^{\varepsilon}}} (x_1x_2x_3x_4x_5x_6)^{-\gamma}\\
&= S_1+S_2+S_3+S_4.
\end{aligned}$$

We write $S_1 \le S_1' + S_1''$ to distinguish the cases $x_1 \le n^{\varepsilon}$ and $x_2 \le n^{\varepsilon}$.

$$\begin{aligned}
S_1' &\ll \sum_{1\le x_1\le n^{\varepsilon}} \sum_{x_2<n} \sum_{x_5} (x_1x_2x_5)^{-\gamma} \sum_{\substack{1\le x_3,x_4\\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma} \sum_{\substack{x_6,x_7\\ x_6+x_7=x_1+x_5}} (x_6x_7)^{-\gamma}\\
&\overset{*}{\ll} \sum_{1\le x_1\le n^{\varepsilon}} \sum_{1\le x_2<n} \sum_{x_5} (x_1x_2x_5)^{-\gamma}(n-x_1-x_2)^{1-2\gamma}(x_1+x_5)^{1-2\gamma}\\
&\overset{*}{\ll} \sum_{1\le x_1\le n^{\varepsilon}} \sum_{1\le x_2<n-x_1} x_1^{2-4\gamma}x_2^{-\gamma}(n-x_1-x_2)^{1-2\gamma} \ll \sum_{1\le x_1\le n^{\varepsilon}} x_1^{2-4\gamma}(n-x_1)^{2-3\gamma}\\
&\overset{*}{\ll} n^{2-3\gamma+\varepsilon(3-4\gamma)} \ll n^{-\frac{\varepsilon}{3+3\varepsilon}+\varepsilon(\frac{1}{3}-\frac{4\varepsilon}{9+9\varepsilon})} \ll n^{-\frac{\varepsilon^2}{9+9\varepsilon}} \ll n^{-\varepsilon^2/18}.
\end{aligned}$$

$$\begin{aligned}
S_1'' &\ll \sum_{1\le x_2\le n^{\varepsilon}} \sum_{1\le x_1<n} \sum_{x_5} (x_1x_2x_5)^{-\gamma} \sum_{\substack{1\varepsilon x_3,x_4\\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma} \sum_{\substack{x_6,x_7\\ x_6+x_7=x_1+x_5}} (x_6x_7)^{-\gamma}\\
&\overset{*}{\ll} \sum_{1\le x_2\le n^{\varepsilon}} \sum_{1\le x_1<n-x_2} \sum_{x_5} (x_1x_2x_5)^{-\gamma}(n-x_1-x_2)^{1-2\gamma}(x_1+x_5)^{1-2\gamma}\\
&\overset{*}{\ll} \sum_{1\le x_2\le n^{\varepsilon}} \sum_{1\le x_1<n-x_2} x_1^{2-4\gamma}x_2^{-\gamma}(n-x_1-x_2)^{1-2\gamma} \overset{*}{\ll} \sum_{1\le x_2\le n^{\varepsilon}} x_2^{-\gamma}(n-x_2)^{4-6\gamma}\\
&\overset{*}{\ll} n^{4-6\gamma+\varepsilon(1-\gamma)} \ll n^{-\frac{2\varepsilon}{3+3\varepsilon}+\varepsilon(\frac{1}{3}-\frac{\varepsilon}{9+9\varepsilon})} \ll n^{\frac{-3\varepsilon+2\varepsilon^2}{9+9\varepsilon}} \ll n^{-\varepsilon^2/18}.
\end{aligned}$$

In the estimates of S_2, S_3 and S_4 we remove the annoying condition $\min(x_1,x_2,x_3,x_4) \le n^{\varepsilon}$.

$$\begin{aligned}
S_2 &\le \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=2x_6}} (x_1x_2x_3x_4x_5x_6)^{-\gamma}\\
&\le \sum_{\substack{1\le x_1,x_2\\ x_1+x_2<n}} (x_1x_2)^{-\gamma}\Big(\sum_{x_5}(x_5(x_1+x_5)/2)^{-\gamma}\Big)\Big(\sum_{\substack{1\le x_3,x_4\\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma}\Big)\\
&\overset{*}{\ll} \sum_{\substack{1\le x_1,x_2\\ x_1+x_2<n}} (x_1x_2)^{-\gamma}x_1^{1-2\gamma}(n-x_1-x_2)^{1-2\gamma} \overset{*}{\ll} \sum_{1\le x_1<n} x_1^{1-3\gamma}(n-x_1)^{2-3\gamma} \ll n^{4-6\gamma}.
\end{aligned}$$

$$
\begin{aligned}
S_3 &\le \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_2=x_6+x_7}} (x_1x_2x_3x_4x_6x_7)^{-\gamma}\\
&\le \sum_{\substack{1\le x_1,x_2\\ x_1+x_2<n}} (x_1x_2)^{-\gamma} \sum_{\substack{1\le x_6,x_7\\ x_6+x_7=x_1+x_2}} (x_6x_7)^{-\gamma} \sum_{\substack{1\le x_3,x_4\\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma}\\
&\overset{*}{\ll} \sum_{\substack{1\le x_1,x_2\\ x_1+x_2<n}} (x_1x_2)^{-\gamma}(x_1+x_2)^{1-2\gamma}(n-x_1-x_2)^{1-2\gamma}\\
&\ll \sum_{1\le l<n} \sum_{\substack{1\le x_1,x_2\\ x_1+x_2=l}} (x_1x_2)^{-\gamma} l^{1-2\gamma}(n-l)^{1-2\gamma} \overset{*}{\ll} \sum_{1\le l<n} l^{2-4\gamma}(n-l)^{1-2\gamma} \overset{*}{\ll} n^{4-6\gamma}.
\end{aligned}
$$

$$
\begin{aligned}
S_4 &\le \sum_{\substack{1\le x_1,x_2,x_3,x_4,x_5,x_6\\ x_1+x_2+x_3+x_4=n\\ x_1+x_5=x_6+x_2}} (x_1x_2x_3x_4x_6x_7)^{-\gamma}\\
&\ll \sum_{\substack{1\le x_1,x_2\\ x_1<x_2\\ x_1+x_2<n}} (x_1x_2)^{-\gamma} \sum_{\substack{1\le x_5,x_6\\ x_5-x_6=x_2-x_1}} (x_6x_5)^{-\gamma} \sum_{\substack{1\le x_3,x_4\\ x_3+x_4=n-x_1-x_2}} (x_3x_4)^{-\gamma}\\
&\overset{*}{\ll} \sum_{\substack{1\le x_1,x_2\\ x_1<x_2\\ x_1+x_2<n}} x_1^{-2\gamma}(x_2-x_1)^{1-2\gamma}(n-x_1-x_2)^{1-2\gamma}\\
&\ll \sum_{1\le x_1<n/2} x_1^{-2\gamma}(n-2x_1)^{3-4\gamma} \ll \sum_{1\le x_1<n/2} x_1^{-2\gamma}(n/2-x_1)^{3-4\gamma} \overset{*}{\ll} n^{3-4\gamma}n^{1-2\gamma}.
\end{aligned}
$$

Thus, $S_2, S_3, S_4 \ll n^{4-6\gamma} \ll n^{-\frac{2\varepsilon}{3+3\varepsilon}} \ll n^{\frac{-\varepsilon^2}{18}}$. □

6.3. *Acknowledgements*

This work has been supported by Project MTM2011-22851 of the MICINN and ICMAT Severo Ochoa project SEV-2011-0087. I am grateful to Igor Shparlinski for useful comments about Theorem 2.2 and to Joel Spencer and Prasad Tetali for providing a copy of the paper [**24**] and for some useful informations about the conjecture of Erdős. I am also grateful to Oriol Serra and Rafa Tesoro for a careful reading of the paper.

References

1. N. Alon, A. Shpilka and C. Umans, 'On Sunflowers and Matrix Multiplication', *Computational Complexity* 22 (2013), 219–243.
2. N. Alon and J. Spencer, 'The Probabilistic Method, Second edition',(Wiley, New York, 2000).
3. J. Cilleruelo, 'Probabilistic constructions of $B_2[g]$ sequences', *Acta Mathematica Sinica* 26 n 7 (2010).
4. J. Cilleruelo, S. Kiss, I. Ruzsa and C. Vinuesa, 'Generalization of a theorem of Erdos and Renyi on Sidon sets', *Random Structures and Algorithms*, 37 n4 (2010).
5. J-M. Deshouillers and A. Plagne, 'A Sidon basis', *Acta Mathematica Hungarica* 123 n3 (2009), 233–238.
6. P. Erdős, 'Einige Bemerkungen zur Arbeit von A. Sthör: Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe', *J. Reine Angew. Math* 197 (1957).
7. P. Erdős, 'The probability method: Successes and limitations', *Journal of Statistical Planning and Inference*, 72 (1998), 207–213.
8. P. Erdős and R. Rado, 'Intersection theorems for systems of sets', *Journal of the London Mathematical Society, Second Series*, 35 (1960), 85–90.
9. P. Erdős and A. Renyi, 'Additive properties of random sequences of positive integers', *Acta Arithmetica*, 6 (1960).

10. P. Erdős, A. Sarkozy and T. Sós, 'On additive properties of general sequences', *Discrete Mathematics* 136, Issue 13, (1994), 75–99.
11. P. Erdős, A. Sarkozy and T. Sós, 'On sum sets of Sidon sets I', *Journal of Number Theory*, 47 (1994), 329–347.
12. P. Erdős and P. Tetali, 'Representations of integers as the sum of k-terms', *Random Structures and Algorithms*, 1 (1990), 245–261.
13. P. Erdős and P. Turán, 'On a problem of Sidon in additive number theory and on some related problems', *Journal of the London Mathematical Society* 16 (1941), 212–215.
14. A. Granville, Shparlinski I.E. and A. Zaharescu, 'On the distribution of rational functions along a curve over $\mathbb{F}p$ and residue races', *J. Number Theory* 112 (2005), 216–237.
15. H. Hasse, 'Zur Theorie der abstrakten elliptischen Funktionenkrper. I, II, III,', *Crelle's Journal* 1936 (175).
16. H. Halberstam and K.F. Roth, 'Sequences', (Clarendon Press, Oxford, 1966).
17. S. Janson, 'Poisson aproximaxion for large desviations', *Randon Structures and Algorithms* 1 (1990), 221-230.
18. J.H. Kim, V.H. Vu 'Concentration of multivariate polynomials and its applications' , *Combinatorica* 20 (2000), 417–434.
19. S. Kiss, 'On generalized Sidon sets which are asymptotic bases'. Preprint.
20. S. Kiss, 'On Sidon sets which are asymptotic basis', *Acta Mathematica Hungarica* 128 (2010), 46–58.
21. S. Kiss, E. Rozgonyi E. and C. Sandor, 'On Sidon sets which are asymptotic basis of order four', *Funct. Approx. Comment. Math.* 51 (2014), no. 2, 393–413.
22. J.H. Kim J.H. and V.H. Vu, 'Concentration of multivariate polynomials and its applications', *Combinatorica* 20 (2000), 417–434.
23. I. Ruzsa, 'Solving a linear equation in a set of integers', *Acta Arithmetica*, LXV.3 (1993) 259–282
24. J. Spencer and P. Tetali, 'Sidon Sets with Small Gaps', *Discrete Probability and Algorithms. The IMA Volumes in Mathematics and its Applications* 72 (1995), 103–109.

Javier Cilleruelo
Instituto de Ciencias Matemáticas (ICMAT) and Departamento de Matemáticas, Universidad Autónoma de Madrid, Madrid 28049
Spain

franciscojavier.cilleruelo@uam.es