COMPOSITIO MATHEMATICA

A solution to the Erdős–Sárközy–Sós problem on asymptotic Sidon bases of order 3

Cédric Pilatte

Compositio Math. **160** (2024), 1418–1432.

doi:10.1112/S0010437X24007140

FOUNDATION COMPOSITIO MATHEMATICA

LONDON MATHEMATICAL SOCIETY
EST. 1865

Compositio Math. 160 (2024) 1418–1432
doi:10.1112/S0010437X24007140

A solution to the Erdős–Sárközy–Sós problem on asymptotic Sidon bases of order 3

Cédric Pilatte

Abstract

A set $S \subset \mathbb{N}$ is a *Sidon set* if all pairwise sums $s_1 + s_2$ (for $s_1, s_2 \in S$, $s_1 \leqslant s_2$) are distinct. A set $S \subset \mathbb{N}$ is an *asymptotic basis of order 3* if every sufficiently large integer n can be written as the sum of three elements of S. In 1993, Erdős, Sárközy and Sós asked whether there exists a set S with both properties. We answer this question in the affirmative. Our proof relies on a deep result of Sawin on the $\mathbb{F}_q[t]$-analogue of Montgomery's conjecture for convolutions of the von Mangoldt function.

1. Introduction

A set S of natural numbers is called a *Sidon set*, or a *Sidon sequence*, if all the sums $s_1 + s_2$, for $s_1, s_2 \in S$, $s_1 \leqslant s_2$, are distinct. Sidon sequences are named after Simon Sidon, who in 1932 asked Erdős about the possible growth rate of such sequences. We write

$$S(x) := \big|\{n \leqslant x : n \in S\}\big|.$$

Erdős observed that the greedy algorithm generates a Sidon sequence S with $S(x) \gg x^{1/3}$. He also conjectured that, for every $\varepsilon > 0$, there is a Sidon sequence S with $S(x) \gg x^{1/2-\varepsilon}$ (see [Erd81]). In some sense, this would be best possible, as Erdős showed that any Sidon sequence S satisfies

$$S(x) \ll x^{1/2}(\log x)^{-1/2}$$

for infinitely many x (see [HR83, Theorem 8]). The lower bound was improved in 1981 by Ajtai, Komlós and Szemerédi [AKS81], who proved the existence of a Sidon sequence S with

$$S(x) \gg x^{1/3} \log x^{1/3}$$

using graph-theoretic tools. In a 1998 landmark paper, Ruzsa [Ruz98] used the fact that the primes form a 'multiplicative Sidon set' to construct a Sidon sequence S with

$$S(x) \gg x^{\sqrt{2}-1-o(1)}.$$

This is still the best-known lower bound: improving the exponent $\sqrt{2} - 1 \approx 0.4142\ldots$ would be a major achievement.

It is worth mentioning that much more is known about *finite* Sidon sets. In particular, the maximal size of a Sidon subset of $\{1, 2, \ldots, n\}$ is $n^{1/2} + O(n^{1/4})$ (see Halberstam and Roth [HR83, Chapter I, § 3] for detailed references).

Received 17 March 2023, accepted in final form 30 November 2023, published online 10 May 2024.
2020 Mathematics Subject Classification 11B13 (primary), 11R58, 11N13, 05D40 (secondary).
Keywords: Sidon sets, additive bases, arithmetic of function fields, probabilistic method.

Sidon sequences have become classical objects of interest in arithmetic combinatorics. It is natural to study the properties of the sumset $S+S := \{s_1+s_2 : s_1, s_2 \in S\}$, or more generally of the kth iterated sumset $kS := \underbrace{S+S+\cdots+S}_{k \text{ times}}$, when S is a Sidon sequence.

A subset $A \subset \mathbb{N}$ is called an *asymptotic basis of order* k if $\mathbb{N} \setminus kA$ is finite. In other words, A is an asymptotic basis of order k if every sufficiently large integer can be written as the sum of k elements of A. Many well-known problems can be restated as follows: for a given set A, what is the smallest k such that A is an asymptotic basis of order k? If A is the set of primes, this is essentially Goldbach's conjecture, whereas if A is the set of perfect dth powers, this is a variant of Waring's problem.

In [ESS94a, Problem 14] and [ESS94b, Problem 8], Erdős, Sárközy and Sós asked the following question.

Problem 1.1. Does there exist a Sidon sequence S which is an asymptotic basis of order 3?

This problem also appears in Erdős [Erd98, p. 212] and Sárközy [Sár01, Problem 32].[1] In his paper [Erd98], Erdős describes Problem 1.1 as 'an old problem of Nathanson and myself'.

It is easy to see that no Sidon sequence can be an asymptotic basis of order 2. In fact, Erdős, Sárközy and Sós [ESS95] proved that for any Sidon set $S \subset \{1, 2, \ldots, n\}$, the sumset $S+S$ cannot contain $\geqslant Cn^{1/2}$ consecutive integers, where C is an absolute constant. Thus, the constant '3' in Problem 1.1 can certainly not be improved.

In this article, we settle Problem 1.1 by proving the existence of a Sidon sequence S which is also an asymptotic basis of order 3.

A number of partial results have been previously obtained in the direction of Problem 1.1.

There is an asymptotic Sidon basis of order	Reference
7	Deshoulliers and Plagne [DP09]
5	Kiss [Kis10]
4	Kiss, Rozgonyi and Sándor [KRS14]
$3+\varepsilon$	Cilleruelo [Cil15, Theorem 1.3]

To make sense of the last row, we recall that a set $S \subset \mathbb{N}$ is an asymptotic basis of order $3+\varepsilon$ if, for every $\varepsilon > 0$, every sufficiently large integer n can be written as the sum of four elements of S, with one of them $\leqslant n^{\varepsilon}$. Our solution to Problem 1.1 also generalises another result of Cilleruelo [Cil15, Theorem 1.2], which states that there is an asymptotic basis S of order 3 such that every integer n can be represented in at most two different ways as a sum of two elements of S. Finally, Kiss and Sándor [KS23] showed the existence of a Sidon sequence $S \subset \mathbb{N}$ whose 3-fold sumset $S+S+S$ has lower asymptotic density > 0.0064.[2]

The results in [Cil15, Kis10, KRS14, KS23] quoted above all use some variant of the probabilistic method. The starting point of all these approaches is to consider a random subset S of $\mathbb{N}$ where each $n \in \mathbb{N}$ is chosen to be in S, independently, with some probability $p(n)$: a typical choice is $p(n) = n^{-\gamma}$ for some $\gamma > 0$. In order for S to resemble a Sidon sequence (i.e. to be able to use the alteration method), it is necessary to have something roughly like $p(n) \ll n^{-2/3}$. In [Cil15, Kis10, KRS14], the authors choose $p(n) = n^{-\gamma}$ for some $\gamma < 2/3$, while for [KS23] they choose $p(n) = cn^{-2/3}$ for some optimised constant $c > 0$. However, for such choices of $p(n)$, the

[1] There is a typo in [Sár01, Problem 32]: '2' should be replaced with '3'.
[2] The constant 0.064 in their paper [KS23] is a typo.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

set S will almost surely *not* be an asymptotic basis of order 3 (this follows from [Gog75, Satz B]). Hence, it seems that purely probabilistic approaches are of little use for addressing Problem 1.1.

In order to obtain an asymptotic basis of order 3, it is more promising to start from the Sidon sequence of Ruzsa mentioned previously [Ruz98], as it is much denser than the Sidon sets obtained by the probabilistic method.

The underlying idea behind Ruzsa's construction is to consider an infinite set of primes $\mathcal{P}$, and for each $p \in \mathcal{P}$, to define a natural number n_p that 'behaves like a logarithm of p', so that an equality $n_{p_1} + n_{p_2} = n_{p_3} + n_{p_4}$ only holds if $p_1 p_2 = p_3 p_4$. Since the primes form a multiplicative Sidon set, we get $\{p_1, p_2\} = \{p_3, p_4\}$, so the set $S := \{n_p : p \in \mathcal{P}\}$ has the Sidon property. In his original paper, Ruzsa defined n_p in terms of the binary expansion of the real number $\log_b p$ (for some $b \in \mathbb{R}^{>1}$) by concatenating certain blocks of digits of $\log_b p$ to obtain a natural number.

Cilleruelo [Cil14] proposed a neat variant of Ruzsa's construction, where the real logarithm is replaced by a family of discrete logarithms. Fix an increasing sequence $\ell_1, \ell_2, \ldots$ of primes $\ell_i \notin \mathcal{P}$ and choose for each i a primitive root ω_i modulo ℓ_i. Write $\log_{\ell_i, \omega_i}(p)$ for the unique $e \in \{0, 1, \ldots, \ell_i - 2\}$ such that $\omega_i^e \equiv p \pmod{\ell_i}$. Cilleruelo defined n_p by encoding many of these discrete logarithms $\log_{\ell_i, \omega_i}(p)$ into a single natural number, using a suitable base expansion. Now, an equality of the form $n_{p_1} + n_{p_2} = n_{p_3} + n_{p_4}$ implies that

$$\log_{\ell_i, \omega_i}(p_1) + \log_{\ell_i, \omega_i}(p_2) = \log_{\ell_i, \omega_i}(p_3) + \log_{\ell_i, \omega_i}(p_4)$$

for many values of i, and thus $p_1 p_2 \equiv p_3 p_4 \pmod{\ell_i}$ for many i. With the appropriate quantification, this is only possible if $p_1 p_2 = p_3 p_4$, which forces $\{p_1, p_2\} = \{p_3, p_4\}$ as before.

Neither Ruzsa's nor Cilleruelo's construction constitutes an asymptotic basis of order 3. There is an obvious obstruction, which is intrinsically tied to the way the information about p is encoded in n_p. For the readers familiar with Ruzsa's construction, the reason is due to the presence of zeros between the blocks of digits of $\log_b p$ within n_p. These zeros cannot be removed, for this would destroy the Sidon property of S. Cilleruelo's construction suffers from a similar drawback.

Our point of departure is Cilleruelo's construction. To avoid the issue raised in the previous paragraph, we replace the zeros in the encoding of n_p with random numbers chosen from a carefully designed set A (see Lemma 3.1 and Definition 3.3). The remaining task is to prove that our modified Cilleruelo construction produces an asymptotic basis of order 3. Eventually, this reduces to information about the distribution of products of three primes in arithmetic progressions.

However, the kind of information we would need is not remotely within reach of our current knowledge of prime numbers. We would need something along the lines of Montgomery's conjecture, a far-reaching generalisation of the generalised Riemann hypothesis. Montgomery's conjecture [IK04, Eq. (17.5), p. 419] states that, for every $\varepsilon > 0$,

$$\psi(x; q, a) := \sum_{\substack{n \leqslant x \\ n \equiv a \pmod{q}}} \Lambda(n) = \frac{x}{\varphi(q)} + O_\varepsilon\big(q^{-1/2} x^{1/2+\varepsilon}\big) \tag{1}$$

uniformly in $x, q \geqslant 1$ and $(a, q) = 1$. This implies the asymptotic formula $\psi(x; q, a) \sim x/\varphi(q)$ uniformly in the residue class and the modulus, provided that $q \leqslant x^{1-\varepsilon}$. We would need a similar (suitably normalised) statement, with the triple convolution $\Lambda * \Lambda * \Lambda$ in place of the von Mangoldt function Λ.

Cilleruelo [Cil14, § 4] observed that his construction works equally well over $\mathbb{F}_q[t]$. Let $\mathcal{F}$ be a set of irreducible monic polynomials in $\mathbb{F}_q[t]$. Fix a sequence (g_i) of irreducible monic polynomials not in $\mathcal{F}$, and for each g_i, choose a generator ω_i of $(\mathbb{F}_q[t]/(g_i))^\times$. For $f \in \mathcal{F}$, write $\log_{g_i, \omega_i}(f)$ for

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

the unique integer $e \in \{0, 1, \dots, q^{\deg g_i} - 1\}$ such that $\omega_i^e \equiv f \pmod{g_i}$. As before, one can define an integer n_f in terms of many such integers $\log_{g_i,\omega_i}(f)$, so that $S := \{n_f : f \in \mathcal{F}\}$ is a Sidon sequence. It is important to remember that S is a Sidon sequence of *integers*, even if $\mathcal{F}$ is a subset of $\mathbb{F}_q[t]$.

The final and crucial ingredient to our proof is the remarkable work of Sawin [Saw21], who proved Montgomery-type results for 'factorisation functions' in $\mathbb{F}_q[t]$, a class of arithmetic functions that encompasses (the $\mathbb{F}_q[t]$-version of) the von Mangoldt function and its convolutions. For example, [Saw21, Theorem 1.2] implies that, for all $\varepsilon > 0$, there is some $q_0(\varepsilon)$ such that, for all prime powers $q \geqslant q_0(\varepsilon)$, the analogue of Montgomery's conjecture (1) holds over $\mathbb{F}_q[t]$ (for squarefree moduli), i.e.

$$\sum_{\substack{f \in \mathbb{F}_q[t] \\ f \text{ monic of degree } n \\ f \equiv a \,(\mathrm{mod}\ g)}} \Lambda(f) = \frac{q^n}{\varphi(g)} + O\big((q^{\deg g})^{-1/2}(q^n)^{1/2+\varepsilon}\big) \tag{2}$$

uniformly in the modulus g (squarefree monic polynomial) and in the residue class a, with $(a, g) = 1$. Sawin's work relies on deep algebraic geometry methods, including sheaf cohomology, the characteristic cycle, vanishing cycles theory and perverse sheaves. We use a bound similar to (2) for the triple convolution $\Lambda * \Lambda * \Lambda$ (see Lemma 5.2). With this powerful result, we are able to prove that our modified Cilleruelo sequence is an asymptotic basis of order 3, thereby solving Problem 1.1.

To conclude this introduction, we mention an application of our work to the study of $B_h[1]$ sequences. A set $S \subset \mathbb{N}$ is said to be $B_h[g]$ set if every positive integer can be written as the sum of h terms from S at most g different ways. For any $h \geqslant 2$, with the probabilistic method, Kiss and Sándor successively showed the existence of a $B_h[1]$ set which is an asymptotic basis of order $2h+1$ (see [KS21]), then $2h$ (see [KS22]). A simple modification of our proof should establish the existence of a $B_h[1]$ set which is an asymptotic basis of order $2h-1$, for $h \geqslant 2$, thus answering a question of Kiss and Sándor [KS21, KS22].

2. Notation

The sumset of two sets $A, B \subset \mathbb{Z}$ is $A + B := \{a + b : a \in A,\ b \in B\}$. We write $f \ll g$ or $f = O(g)$ if $|f| \leqslant Cg$ for some absolute constant $C > 0$. If instead C depends on a parameter θ, we write $f \ll_\theta g$ or $f = O_\theta(g)$.

For $b \geqslant 1$ and $a \in \mathbb{Z}$, we write $[a \bmod b]$ for the unique integer n satisfying $0 \leqslant n < b$ and $n \equiv a \pmod{b}$.

Definition 2.1 (Generalised base). Let $\boldsymbol{b} = (b_1, b_2, \dots)$ be an infinite sequence of integers $\geqslant 2$. For any $n \geqslant 1$ and any $x_1, \dots, x_n \in \mathbb{Z}$, we write

$$\overline{x_n \dots x_1}^{\boldsymbol{b}} := x_1 + x_2 b_1 + x_3 b_1 b_2 + \cdots + x_n b_1 b_2 \cdots b_{n-1}.$$

Any $m \in \mathbb{Z}^{\geqslant 1}$ can be uniquely represented as $m = \overline{x_n \dots x_1}^{\boldsymbol{b}}$ for some $0 \leqslant x_i < b_i$ with $x_n \neq 0$. Recall, however, that the notation $\overline{x_n \dots x_1}^{\boldsymbol{b}}$ is defined for arbitrary integers x_i: we do not always require $0 \leqslant x_i < b_i$.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

3. Construction

The following lemma is a slight strengthening of the statement that, for any sufficiently large p, there exists a set $A \subset \mathbb{Z}/p\mathbb{Z}$ such that A and $A+A$ are disjoint, and moreover $A+A+A = \mathbb{Z}/p\mathbb{Z}$.

LEMMA 3.1. *For every sufficiently large prime p, there is a set $A \subset \{1, 2, \ldots, \lfloor p/2 \rfloor - 1\}$ such that:*

(i) *the sets A and $A + A + \{0, 1\}$ are disjoint;*
(ii) *$A + A + A$ contains $p + 2$ consecutive integers.*

It turns out that we only need to use the existence of a single pair (p, A) with these properties. Lemma 3.1 can be shown by the alteration method in probabilistic combinatorics. We provide a detailed proof in Appendix A.

For the remainder of this paper, we fix a pair (p, A) satisfying the conclusion of Lemma 3.1. In particular, p should be thought of as an absolute constant. Let $c = 0.35$, the point is that

$$\frac{1}{3} < c < \frac{3 - \sqrt{5}}{2}.$$

Let $C_0 > 100p$ be a large absolute constant that will be chosen later.

DEFINITION 3.2 ($\mathcal{P}_d$, g_i, ω_i, $\mathcal{F}_k$, $\mathcal{F}$). Let $q \geqslant C_0$ be a prime (or a prime power). Let $\mathcal{P}_d$ be the set of irreducible monic polynomials $f \in \mathbb{F}_q[t]$ of degree d. Recall the standard formula of Gauss

$$|\mathcal{P}_d| = \frac{1}{d} \sum_{e \mid d} \mu\left(\frac{d}{e}\right) q^e = \frac{q^d}{d} + O(q^{d/2}). \tag{3}$$

For $i \geqslant 1$, let g_i be an arbitrary element of $\mathcal{P}_{2i-1}$, and fix an arbitrary generator ω_i of $(\mathbb{F}_q[t]/(g_i))^\times$.

For $k \geqslant C_0$, let

$$\mathcal{F}_k := \bigcup_{ck^2 \leqslant 2i < c(k+1)^2} \mathcal{P}_{2i}.$$

Let $\mathcal{F} := \bigcup_{k \geqslant C_0} \mathcal{F}_k$.

We work in the generalised base $\boldsymbol{b} = (b_1, b_2, \ldots)$ where

$$b_i = \begin{cases} q^i - 1 & i \text{ odd}, \\ p & i \text{ even}. \end{cases} \tag{4}$$

DEFINITION 3.3 (e_i, r_i, s, n_f, S). Let $k \geqslant C_0$ and let $f \in \mathcal{F}_k$. We associate to f a positive integer n_f as follows.

- For $1 \leqslant i \leqslant k$, let $e_i(f)$ be the unique integer such that $0 \leqslant e_i(f) < b_{2i-1} = q^{2i-1} - 1$ and
$$\omega_i^{e_i(f)} \equiv f \pmod{g_i}.$$
- Let $r_1(f), \ldots, r_k(f)$ be a sequence of independent and identically distributed (i.i.d.) random variables, each uniformly distributed on A (in particular, $1 \leqslant r_i(f) \leqslant p/2 - 1$).
- Let $s(f)$ be an integer chosen uniformly at random in the set $\{1, 2, \ldots, q^{3k}\}$.

It is understood that the family of all random variables $r_i(f)$ and $s(f)$ (over all choices of f and i) is independent.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

We define n_f to be the integer

$$n_f := \overline{s\, r_k\, e_k\, \dots\, r_2\, e_2\, r_1\, e_1}^{\boldsymbol{b}}$$

omitting the dependence of each digit on f for conciseness. To be explicit,

$$n_f = e_1(f) + r_1(f)b_1 + e_2(f)b_1b_2 + r_2(f)b_1b_2b_3 + \cdots + r_k(f)b_1b_2\cdots b_{2k-1} + s(f)b_1b_2\cdots b_{2k},$$

with b_i as in (4).[3]

Finally, we define $S = \{n_f : f \in \mathcal{F}\}$.

We show that, with probability 1, the elements of S form a Sidon sequence and an asymptotic basis of order 3.

Lemma 3.4.

(i) *Let $f \in \mathcal{F}_k$. Then $n_f = q^{k^2+O(k)}$.*

(ii) *Let $f_1, f_2 \in \mathcal{F}$ be such that $n_{f_1} = n_{f_2}$. Then $f_1 = f_2$.*

Proof. (i) On the one hand,

$$n_f = \overline{s\, r_k\, e_k\, \dots\, r_2\, e_2\, r_1\, e_1}^{\boldsymbol{b}} \geqslant \overline{1\, \underbrace{0\,0\, \dots\, 0\,0}_{2k \text{ zeros}}}^{\boldsymbol{b}} = \prod_{i=1}^{k} p(q^{2i-1}-1) > \prod_{i=1}^{k} q^{2i-1} = q^{k^2}.$$

On the other hand,

$$n_f = \overline{s\, r_k\, e_k\, \dots\, r_2\, e_2\, r_1\, e_1}^{\boldsymbol{b}} \leqslant \overline{(q^{3k}+1)\, \underbrace{0\,0\, \dots\, 0\,0}_{2k \text{ zeros}}}^{\boldsymbol{b}} = (q^{3k}+1)\prod_{i=1}^{k} p(q^{2i-1}-1) = q^{k^2+O(k)},$$

as $p \leqslant q$. This shows that $n_f = q^{k^2+O(k)}$.

(ii) Let $n = n_{f_1} = n_{f_2}$. Suppose that $f_1 \in \mathcal{F}_{k_1}$ and $f_2 \in \mathcal{F}_{k_2}$. By part (i), we have

$$q^{k_1^2+O(k_1)} = n = q^{k_2^2+O(k_2)},$$

from which we infer that $k_1 = k_2 + O(1)$. Let $k = \min(k_1, k_2)$. Since $n_{f_1} = n_{f_2} = n$, we also have

$$\overline{r_k(f_1)\, e_k(f_1)\, \dots\, r_1(f_1)\, e_1(f_1)}^{\boldsymbol{b}} = \left[n \bmod b_1b_2\cdots b_{2k}\right] = \overline{r_k(f_2)\, e_k(f_2)\, \dots\, r_1(f_2)\, e_1(f_2)}^{\boldsymbol{b}}.$$

The left- and right-hand side are two base-$\boldsymbol{b}$ representations of the same natural number, and since the ith digit is between 0 and $b_i - 1$ in both cases, we see that $e_i(f_1) = e_i(f_2)$ (and $r_i(f_1) = r_i(f_2)$) for all $1 \leqslant i \leqslant k$. By the definition of e_i, this implies that $f_1 \equiv f_2 \pmod{g_i}$. By the Chinese remainder theorem, we deduce that

$$f_1 \equiv f_2 \pmod{g_1 \cdots g_k}. \tag{5}$$

However, $\deg(f_1 - f_2) \leqslant ck^2 + O(k)$ by definition of $\mathcal{F}_k$, whereas $\deg(g_1 \cdots g_k) = k^2$. Therefore, $\deg(f_1 - f_2) < \deg(g_1 \cdots g_k)$ if C_0 is sufficiently large, and (5) can only hold if $f_1 = f_2$. □

4. Sidon property

Lemma 4.1. *Let $f_1, f_2, f_3, f_4 \in \mathcal{F}$ be such that $n_{f_1} + n_{f_2} = n_{f_3} + n_{f_4}$. Then $\{f_1, f_2\} = \{f_3, f_4\}$.*

Proof. Without loss of generality, suppose that $f_i \in \mathcal{F}_{k_i}$ for $i \in \{1, 2, 3, 4\}$, where $k_1 \geqslant k_2$ and $k_3 \geqslant k_4$. By part (i) of Lemma 3.4, we have $q^{k_1^2+O(k_1)} = n_{f_1} + n_{f_2} = n_{f_3} + n_{f_4} = q^{k_3^2+O(k_3)}$ and, thus, $k_1 = k_3 + O(1)$.

[3] Note that $s(f)$ may exceed b_{2k+1}, whereas for $1 \leqslant i \leqslant 2k$ the ith digit ($e_{(i+1)/2}$ or $r_{i/2}$) is always between 0 and $b_i - 1$.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

We claim that $k_2 = k_4 + O(1)$. Write the integer $n := n_{f_1} + n_{f_2} = n_{f_3} + n_{f_4}$ in base $\boldsymbol{b}$, as

$$n = \overline{y_l\, x_l\, \ldots\, y_2\, x_2\, y_1\, x_1}^{\boldsymbol{b}},$$

for some $0 \leqslant x_i < q^{2i-1} - 1$ and $0 \leqslant y_1 < p$, where x_l and y_l are not both zero. Let us align the base-$\boldsymbol{b}$ digits of n_{f_1}, n_{f_2} and n:

n_{f_1} :	$s(f_1)$	$r_{k_1}(f_1)$	$e_{k_1}(f_1)$	$\cdots$	$r_{k_2+1}(f_1)$	$e_{k_2+1}(f_1)$	$r_{k_2}(f_1)$	$e_{k_2}(f_1)$	$\cdots$	$r_1(f_1)$	$e_1(f_1)$
n_{f_2} :						$s(f_2)$	$r_{k_2}(f_2)$	$e_{k_2}(f_2)$	$\cdots$	$r_1(f_2)$	$e_1(f_2)$
n :	$\cdots$	y_{k_1}	x_{k_1}	$\cdots$	y_{k_2+1}	x_{k_2+1}	y_{k_2}	x_{k_2}	$\cdots$	y_1	x_1

Since $n_{f_1} + n_{f_2} = n$ and $r_i(f_j) \leqslant p/2 - 1$ for all i and j, we can easily express the digits x_i and y_i in terms of the digits of n_{f_1} and n_{f_2}. The computation gets slightly more complicated when i is close to k_2 (or larger than k_1) because $s(f_2)$ can be as large as q^{3k_2}, which exceeds $b_{2k_2+1} = q^{2k_2+1} - 1$.

On the one hand, for $1 \leqslant i \leqslant k_2$, we have:

- $x_i = \left[e_i(f_1) + e_i(f_2) \bmod q^{2i-1} - 1\right]$; and
- $y_i = r_i(f_1) + r_i(f_2) + \mathbf{1}_{e_i(f_1)+e_i(f_2) \geqslant q^{2i-1}-1}$.

On the other hand, for $k_2 + 3 \leqslant i \leqslant k_1$ we have:[4]

- $x_i = e_i(f_1)$; and
- $y_i = r_i(f_1)$.

In particular, for $1 \leqslant i \leqslant k_2$ we have $y_i \in A + A + \{0,1\}$, whereas for $k_2 + 3 \leqslant i \leqslant k_1$ we have $y_i \in A$. Let i_0 be the largest integer $\leqslant l$ such that $y_1, \ldots, y_{i_0} \in A + A + \{0,1\}$. If $k_2 + 3 > k_1$, then $i_0 = k_1 + O(1) = k_2 + O(1)$. Otherwise, since A and $A + A + \{0,1\}$ are disjoint by Lemma 3.1(i), we have $k_2 \leqslant i_0 < k_2 + 3$. Hence, $i_0 = k_2 + O(1)$ in both cases.

Repeating the whole argument with f_3 and f_4 in place of f_1 and f_2, we obtain that $i_0 = k_4 + O(1)$, whence $k_2 = k_4 + O(1)$ as claimed.

Our previous computations have shown that

$$x_i = \left[e_i(f_1) + e_i(f_2) \bmod q^{2i-1} - 1\right] = \left[e_i(f_3) + e_i(f_4) \bmod q^{2i-1} - 1\right]$$

for $1 \leqslant i \leqslant \min(k_2, k_4)$, and

$$x_i = e_i(f_1) = e_i(f_3)$$

for $\max(k_2, k_4) + 3 \leqslant i \leqslant \min(k_1, k_3)$.

By the definition of $e_i(f)$ and the Chinese remainder theorem, this implies that

$$f_1 f_2 \equiv f_3 f_4 \left(\bmod \prod_{i=1}^{\min(k_2,k_4)} g_i\right) \tag{6}$$

and

$$f_1 \equiv f_3 \left(\bmod \prod_{i=\max(k_2,k_4)+3}^{\min(k_1,k_3)} g_i\right). \tag{7}$$

Suppose first that $f_1 = f_3$. Then $n_{f_2} = n - n_{f_1} = n - n_{f_3} = n_{f_4}$ and, hence, $f_2 = f_4$ by Lemma 3.4(ii), which is what we needed to prove.

[4] Note that $s(f_2) \leqslant q^{3k_2} < b_{2k_2+1} b_{2k_2+2} b_{2k_2+3}$, which is why these formulas are valid for $k_2 + 3 \leqslant i \leqslant k_1$.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

If $f_1 f_2 = f_3 f_4$, we immediately conclude that $\{f_1, f_2\} = \{f_3, f_4\}$, since all f_i are irreducible monic polynomials.

Hence, we may suppose that $f_1 \neq f_3$ and $f_1 f_2 \neq f_3 f_4$. By (7) we must have

$$ck_1^2 + O(k_1) \geqslant \deg(f_1 - f_3) \geqslant \deg\left(\prod_{i=k_2+O(1)}^{k_1-O(1)} g_i\right) = k_1^2 - k_2^2 - O(k_1).$$

Similarly, by (6), we have

$$ck_1^2 + ck_2^2 + O(k_1) \geqslant \deg(f_1 f_2 - f_3 f_4) \geqslant \deg\left(\prod_{i=1}^{k_2-O(1)} g_i\right) = k_2^2 - O(k_2).$$

Therefore, we have

$$\begin{cases} k_1^2 - k_2^2 \leqslant ck_1^2 + O(k_1), \\ k_2^2 \leqslant ck_1^2 + ck_2^2 + O(k_1). \end{cases}$$

Hence,

$$(1-c)k_1^2 - O(k_1) \leqslant k_2^2 \leqslant \frac{c}{1-c}k_1^2 + O(k_1),$$

which is impossible since $1 - c > c/(1-c)$, if C_0 is sufficiently large. □

5. Asymptotic basis of order 3

Lemma 5.1. *Let $m \geqslant 3$ be an integer. We can write $m = \overline{z\, y_k\, x_k\, \ldots\, y_2\, x_2\, y_1\, x_1}^{\boldsymbol{b}}$, for some $k \geqslant 0$ and some integers x_i, y_i, z satisfying:*

- $0 \leqslant x_i < q^{2i-1} - 1$ *for all* $1 \leqslant i \leqslant k$*;*
- $0 \leqslant y_i < 2p$ *for all* $1 \leqslant i \leqslant k$*;*
- $\{y_i - 2, y_i - 1, y_i\} \subset A + A + A$ *for all* $1 \leqslant i \leqslant k$*;*
- $3 \leqslant z \leqslant 6pq^{2k+1}$.

Proof. We proceed in a similar way to the standard algorithm that generates the base-b expansion of an integer. Let us inductively define a sequence (m_i) of integers $\geqslant 3$. Let $m_1 := m$. Suppose we have defined $m_1, \ldots, m_{2l-1}$ for some $l \geqslant 1$.

If $m_{2l-1} > 6pq^{2l-1}$, we define

$$x_l := \left[m_{2l-1} \bmod q^{2l-1} - 1\right]$$

and we set $m_{2l} := (m_{2l-1} - x_l)/(q^{2l-1} - 1)$.

Let $0 \leqslant y_l < 2p$ be an integer with $y_l \equiv m_{2l} \pmod p$ and such that

$$\{y_l - 2, y_l - 1, y_l\} \subset A + A + A.$$

Such an integer exists by Lemma 3.1(ii) (note that $A + A + A \subset [0, 3p/2]$). We define $m_{2l+1} := (m_{2l} - y_l)/p$. Since $m_{2l-1} > 6pq^{2l-1}$, we have

$$m_{2l+1} \geqslant \frac{m_{2l}}{p} - 2 \geqslant \frac{m_{2l-1}}{pq^{2l-1}} - 3 \geqslant 3.$$

We have thus constructed x_l, y_l, m_{2l} and m_{2l+1}.

Otherwise, if $m_{2l-1} \leqslant 6pq^{2l-1}$, so we can define $z := m_{2l-1}$ and stop the construction (thus, $k = l - 1$).

The procedure must stop since (m_i) is decreasing. When the procedure terminates, we end up with integers $x_1, \ldots, x_k$, $y_1, \ldots, y_k$ and z that satisfy the required properties, by construction. □

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

Lemma 5.2. *Let $d \geqslant 1$ and let $g \in \mathbb{F}_q[t]$ be a squarefree monic polynomial of degree*

$$2 \leqslant \deg(g) \leqslant 3\theta d$$

for some $0 < \theta < 1$. Let $a \in (\mathbb{F}_q[t]/(g))^\times$. Then, with $\phi(g) := \left|(\mathbb{F}_q[t]/(g))^\times\right|$, we have

$$\left| \sum_{\substack{f_1,f_2,f_3 \in \mathcal{P}_d \\ f_i \text{ distinct} \\ f_1f_2f_3 \equiv a \pmod{g}}} 1 - \frac{1}{\phi(g)} \binom{|\mathcal{P}_d|}{3} \right| \ll e^{O_\theta(d)} q^{(3d-\deg(g))/2}.$$

Proof. We use [Saw21, Lemma 9.14] with $\omega = 3$, $n_1 = n_2 = n_3 = d$, $n = 3d$ and $h = 1$ (meaning that $c = 0$). The author writes $\mathbb{F}_q[t]^+$ and $\mathcal{M}_n$ for the set of monic polynomials in $\mathbb{F}_q[t]$ and the set of monic polynomials of degree n in $\mathbb{F}_q[t]$, respectively. The definition of $H^h_{n_1,\dots,n_\omega}(f)$ can be found at the bottom of [Saw21, p. 89]. In particular,

$$H^1_{d,d,d}(f) = d^3 \sum_{\substack{f_1,f_2,f_3 \in \mathcal{P}_d \\ f_i \text{ distinct} \\ f_1f_2f_3 = f}} 1,$$

and, thus,

$$\sum_{f \in \mathcal{M}_{3d}} H^1_{d,d,d}(f) = d^3 \binom{|\mathcal{P}_d|}{3}.$$

Since $c = 0$, the conclusion of [Saw21, Lemma 9.14] simplifies to

$$d^3 \left| \sum_{\substack{f_1,f_2,f_3 \in \mathcal{P}_d \\ f_i \text{ distinct} \\ f_1f_2f_3 \equiv a (\mathrm{mod}(g))}} 1 - \frac{1}{\phi(g)} \binom{|\mathcal{P}_d|}{3} \right| \ll \frac{(3d)!}{(d!)^3} (O_\theta(q))^{(3d-\deg(g))/2}.$$

By Stirling's formula, $(3d)!/(d!)^3 \ll e^{O(d)}$, and we obtain Lemma 5.2. □

Lemma 5.3. *Let $m \geqslant q^{(C_0+2)^2}$. Then $\mathbb{P}(m \notin S + S + S) \leqslant \exp(-m^{3c-1-o(1)})$ as $m \to +\infty$.*[5]

Proof. Let $m \geqslant q^{(C_0+2)^2}$. We write $m = \overline{z\, y_k\, x_k\, \dots\, y_2\, x_2\, y_1\, x_1}^{\boldsymbol{b}}$ with x_i, y_i and z satisfying the conclusion of Lemma 5.1. By a computation similar to that in Lemma 3.4(i), we have

$$q^{k^2} \leqslant m \leqslant q^{(k+2)^2}.$$

In particular, $k \geqslant C_0$ and $m = q^{k^2+O(k)}$.

We show that m is very likely to be expressible as $n_{f_1} + n_{f_2} + n_{f_3}$ for some $f_i \in \mathcal{F}_k$.

We first focus on the digits x_i. By the Chinese remainder theorem, there is some $a \in \mathbb{F}_q[t]$ such that

$$a \equiv \omega_i^{x_i} \pmod{g_i}$$

for all $1 \leqslant i \leqslant k$. Let $g = \prod_{1 \leqslant i \leqslant k} g_i$; it is a squarefree polynomial of degree $\deg(g) = k^2$. Let d be an even integer with $ck^2 \leqslant d < c(k+1)^2$. Let $\mathcal{E}$ be the set of all triples $(f_1, f_2, f_3) \in (\mathcal{P}_d)^3$ of distinct polynomials such that $f_1f_2f_3 \equiv a \pmod{g}$. By Lemma 5.2 with $\theta = k^2/(3d) \leqslant 1/(3c) < 1$, we have

$$\left| |\mathcal{E}| - \frac{1}{\phi(g)} \binom{|\mathcal{P}_d|}{3} \right| \ll e^{O(d)} q^{(3d-k^2)/2}. \tag{8}$$

[5] Recall that $\mathbb{P}$ refers to the probability in the random choice of $r_i(f), s(f)$ (which are the random variables used to define S).

Using the fact that $\phi(g) = \prod_{i=1}^{k}(q^{2i-1}-1) = q^{k^2 - O(k)}$ and the bound (3) on the size of $\mathcal{P}_d$, we can simplify (8) to get

$$\left|\mathcal{E}\right| = q^{3d-k^2+O(k+\log d)} + O\left(q^{(3d-k^2)/2+O(d/\log q)}\right) = q^{(3c-1)k^2+O(k)} + O\left(q^{((3c-1)/2+O(1/\log q))k^2}\right),$$

using $d = ck^2 + O(k)$ for the last step. If C_0 is sufficiently large, the error term is negligible and we obtain

$$\left|\mathcal{E}\right| = q^{(3c-1)k^2+O(k)}. \tag{9}$$

Let us estimate, for a fixed triple $(f_1, f_2, f_3) \in \mathcal{E}$, the probability that $n_{f_1} + n_{f_2} + n_{f_3} = m$. By the definition of n_{f_i}, we can write

$$n_{f_1} + n_{f_2} + n_{f_3} = \overline{s'\, r'_k\, e'_k\, \ldots\, r'_2\, e'_2\, r'_1\, e'_1}^{\,\boldsymbol{b}}$$

where e'_i, r'_i and s' are defined by:

- $e'_i = \left[e_i(f_1) + e_i(f_2) + e_i(f_3) \bmod q^{2i-1}-1\right]$ for $1 \leqslant i \leqslant k$;
- $r'_i = r_i(f_1) + r_i(f_2) + r_i(f_3) + \kappa_i$ for $1 \leqslant i \leqslant k$, with

$$\kappa_i := \left\lfloor \frac{e_i(f_1) + e_i(f_2) + e_i(f_3)}{q^{2i-1}-1} \right\rfloor \in \{0,1,2\};$$

- $s' = s(f_1) + s(f_2) + s(f_3)$.

The assumption that $(f_1, f_2, f_3) \in \mathcal{E}$ ensures that $e'_i = x_i$ for all $1 \leqslant i \leqslant k$. Indeed, both e'_i and x_i are between 0 and $q^{2i-1} - 2$, and

$$\omega_i^{e'_i} \equiv \omega_i^{e_i(f_1)+e_i(f_2)+e_i(f_3)} \equiv f_1 f_2 f_3 \equiv a \equiv \omega_i^{x_i} \pmod{g_i}.$$

Hence,

$$\mathbb{P}\left(n_{f_1} + n_{f_2} + n_{f_3} = m\right) \geqslant \mathbb{P}\left(r'_1 = y_1,\, r'_2 = y_2,\, \ldots, r'_k = y_k,\, s' = z\right) = \mathbb{P}\left(s' = z\right) \prod_{1 \leqslant i \leqslant k} \mathbb{P}\left(r'_i = y_i\right),$$

where we used the independence of the family of all random variables $r_i(f_j)$ and $s(f_j)$ in the last step.

Recall that s' is the sum of three independent random numbers chosen uniformly in $\{1, 2, \ldots, q^{3k}\}$, so s' can be equal to any integer in the set $\{3, 4, \ldots, 3q^{3k}\}$, each of which occurs with probability at least q^{-9k}. Since z is a fixed element of $\{3, 4, \ldots, 6pq^{2k+1}\}$ and $6pq^{2k+1} \leqslant 3q^{3k}$, we get

$$\mathbb{P}\left(s' = z\right) \geqslant q^{-9k}.$$

Similarly, r'_i is the sum of three elements of A chosen uniformly and independently at random, plus a fixed carry $\kappa_i \in \{0,1,2\}$. Since $\{y_i - 2, y_i - 1, y_i\} \subset A + A + A$ by Lemma 5.1, there is at least one triple $(a_1, a_2, a_3) \in A^3$ such that $a_1 + a_2 + a_3 + \kappa_i = y_i$. Thus, we see that $\mathbb{P}\left(r'_i = y_i\right) \geqslant |A|^{-3} \geqslant p^{-3}$. We have thus shown that

$$\mathbb{P}\left(n_{f_1} + n_{f_2} + n_{f_3} = m\right) \geqslant q^{-9k} p^{-3k} \geqslant q^{-O(k)}.$$

To conclude the argument, we wish to restrict to a large subset $\mathcal{E}_0$ of $\mathcal{E}$ such that no polynomial appears in more than one triple of $\mathcal{E}_0$. Observe that, if two triples (f, f_1, f_2) and (f, f_3, f_4) are both in $\mathcal{E}$, then $\{f_1, f_2\} = \{f_3, f_4\}$. Indeed, two such triples satisfy $f_1 f_2 \equiv f^{-1} a \equiv f_3 f_4 \pmod{g}$,

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

which implies that $f_1f_2 = f_3f_4$ in $\mathbb{F}_q[t]$ since

$$\deg(f_1f_2 - f_3f_4) \leqslant 2d < k^2 = \deg(g),$$

and thus $\{f_1, f_2\} = \{f_3, f_4\}$. In addition, if $(f_1, f_2, f_3) \in \mathcal{E}$ then so does $(f_{\sigma(1)}, f_{\sigma(2)}, f_{\sigma(3)})$ for every permutation σ. Therefore, any given polynomial $f \in \mathcal{P}_d$ appears at most once in a triple of $\mathcal{E}$, up to permutations. This shows that there is a suitable subset $\mathcal{E}_0$ with $\left|\mathcal{E}_0\right| = \frac{1}{6}\left|\mathcal{E}\right|$.

Therefore,

$$\begin{aligned}\mathbb{P}\big(m \notin S+S+S\big) &\leqslant \mathbb{P}\Bigg(\bigcap_{(f_1,f_2,f_3)\in\mathcal{E}_0} \{n_{f_1}+n_{f_2}+n_{f_3} \neq m\}\Bigg)\\ &= \prod_{(f_1,f_2,f_3)\in\mathcal{E}_0} \mathbb{P}\big(n_{f_1}+n_{f_2}+n_{f_3} \neq m\big) \leqslant \big(1-q^{-O(k)}\big)^{|\mathcal{E}_0|},\end{aligned}$$

using the independence of the family of random variables $(n_{f_1}+n_{f_2}+n_{f_3})_{(f_1,f_2,f_3)\in\mathcal{E}_0}$. By (9) and the simple inequality $1-x \leqslant \exp(-x)$, we get

$$\mathbb{P}\big(m \notin S+S+S\big) \leqslant \exp\left(-q^{-O(k)}q^{(3c-1)k^2-O(k)}\right) \leqslant \exp\left(-m^{(3c-1+o(1))}\right),$$

as $m = q^{k^2+O(k)}$. □

Theorem 5.4. *With probability* 1, *the elements of* S *form a Sidon sequence and an asymptotic basis of order* 3.

Proof. By Lemma 4.1, S always has the Sidon property. By Lemma 5.3,

$$\mathbb{P}\big(m \notin S+S+S\big) \leqslant \exp(-m^{3c-1-o(1)}) \ll m^{-2}$$

for sufficiently large m, so

$$\mathbb{P}\big(\left|\mathbb{N} \setminus S+S+S\right| = +\infty\big) = 0$$

by the Borel–Cantelli lemma. We conclude that S is almost surely an asymptotic basis of order 3. □

Acknowledgements

I would like to thank my advisors, Ben Green and James Maynard, for their expert guidance and continuing encouragement. I am very grateful to Oliver Riordan for pointing out an error in the appendix of an earlier version of this paper, and for suggesting the method used here to fix it. The author is supported by the Oxford Mathematical Institute and a Saven European Scholarship.

Conflicts of Interest

None.

Appendix A. Auxiliary set

In this section, we use the notation $f * g$ for the additive convolution

$$f*g(n) = (f*g)(n) := \sum_{\substack{a,b\in\mathbb{Z}\\ a+b=n}} f(a)g(b).$$

Proof of Lemma 3.1. Let $I = \{1, 2, \ldots, \lfloor p/2 \rfloor - 1\} \subset \mathbb{Z}$.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

Let R be the random set obtained by selecting each element of I, independently, with probability $Kp^{-2/3}$, where $K = \lceil \log p \rceil$. In particular, $\mathbb{E}\big[|R|\big] \asymp Kp^{1/3}$.

Note that R almost satisfies property (i) of Lemma 3.1, in the sense that

$$\mathbb{E}\big[\big|R \cap (R+R+\{0,1\})\big|\big] \leqslant \sum_{\substack{a,b,c\in I\\ a-b-c\in\{0,1\}}} \mathbb{P}\big(a,b,c \in R\big).$$

The probability $\mathbb{P}\big(a,b,c\in R\big)$ is $(Kp^{-2/3})^k$ where $k = |\{a,b,c\}|$. Therefore,

$$\mathbb{E}\big[\big|R \cap (R+R+\{0,1\})\big|\big] \ll p^2(Kp^{-2/3})^3 + p(Kp^{-2/3})^2 + 1(Kp^{-2/3})^1 \ll K^3. \tag{A.1}$$

It will be useful to know the bound $\big\|\mathbf{1}_R * \mathbf{1}_R\big\|_\infty \ll 1$, which holds with high probability. Indeed, if $\mathbf{1}_R * \mathbf{1}_R(n) \geqslant 9$ for some $n\in\mathbb{N}$, then R contains elements $a_1 < a_2 < \cdots < a_8$ such that $a_i + a_{9-i} = n$ for all i and, thus,

$$\mathbb{P}\big(\exists n,\ \mathbf{1}_R * \mathbf{1}_R(n) \geqslant 9\big) \leqslant \sum_{n=1}^{p} \sum_{\substack{a_1,\dots,a_8\in I\\ a_i \text{ distinct}\\ a_i+a_{9-i}=n}} \mathbb{P}\big(\forall i,\ a_i\in R\big) \ll p^5(Kp^{-2/3})^8 \ll K^8p^{-1/3} \ll K^{-1}. \tag{A.2}$$

Similarly, if $\mathbf{1}_R * \mathbf{1}_{-R}(n) \geqslant 8$ for some $n\in\mathbb{Z}^{\neq 0}$, then R contains distinct elements $a_1,\dots,a_8$ such that $a_{2i} - a_{2i-1} = n$ for $i\in\{1,2,3,4\}$, and as before we get

$$\mathbb{P}\big(\exists n\in\mathbb{Z}^{\neq 0},\ \mathbf{1}_R * \mathbf{1}_{-R}(n) \geqslant 8\big) \ll K^{-1}. \tag{A.3}$$

We now turn to property (ii) of Lemma 3.1. Let n be an integer with $p/5 \leqslant n \leqslant 7p/5$. We show that n can be written in relatively many ways as the sum of three distinct elements of R, with high probability. Let $\mathcal{T}(n)$ be the collection of all sets $\{a,b,c\}$ of three pairwise distinct elements of I such that $a+b+c=n$. We show that

$$\mathbb{P}\Bigg(\sum_{T\in\mathcal{T}(n)} \mathbf{1}_{T\subset R} \leqslant K^2\Bigg) \ll e^{-K^2}. \tag{A.4}$$

To bound the left-hand side, we use Janson's inequality. This inequality involves the quantities

$$\lambda := \mathbb{E}\Bigg[\sum_{T\in\mathcal{T}(n)} \mathbf{1}_{T\subset R}\Bigg] = \sum_{T\in\mathcal{T}(n)} (Kp^{-2/3})^3 \gg p^2(Kp^{-2/3})^3 = K^3$$

and

$$\Delta := \sum_{\substack{T_1,T_2\in\mathcal{T}(n)\\ T_1\cap T_2\neq\emptyset\\ T_1\neq T_2}} \mathbb{P}\big(T_1\cup T_2 \subset R\big).$$

Janson's inequality [JRL11, Theorem 2.14] gives

$$\mathbb{P}\Bigg(\sum_{T\in\mathcal{T}(n)} \mathbf{1}_{T\subset R} \leqslant \tfrac{1}{2}\lambda\Bigg) \leqslant \exp\left(-\frac{\lambda^2}{8(\lambda+\Delta)}\right). \tag{A.5}$$

Observe that, if $T_1, T_2 \in \mathcal{T}(n)$ are distinct but not disjoint, we must have $\big|T_1\cap T_2\big| = 1$, so there are pairwise distinct elements $a, b_1, b_2, c_1, c_2 \in I$ such that $T_1 = \{a, b_1, c_1\}$ and

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

$T_2 = \{a, b_2, c_2\}$. In particular,

$$\mathbb{P}\big(T_1 \cup T_2 \subset R\big) = (Kp^{-2/3})^5,$$

and there are $\ll p^3$ such pairs (T_1, T_2). Hence,

$$\Delta \ll p^3 (Kp^{-2/3})^5 \ll 1 \ll \lambda.$$

Provided p is larger than some absolute constant, we have $K^2 \leqslant \frac{1}{2}\lambda$, so (A.4) follows from (A.5), using our estimates for λ and Δ. By the union bound, we get

$$\mathbb{P}\Bigg(\exists n \in [p/5, 7p/5] \cap \mathbb{Z}, \sum_{T \in \mathcal{T}(n)} \mathbf{1}_{T \subset R} \leqslant K^2\Bigg) \ll pe^{-K^2} \ll K^{-1}. \tag{A.6}$$

We have shown that, typically, R satisfies property (ii) of Lemma 3.1 in a robust sense (by (A.6)), but R just fails to satisfy property (i) (see (A.1)).

We define $X := R \cap (R + R + \{0, 1\})$ and $A := R \setminus X$. This set A satisfies property (i) of Lemma 3.1 by construction. We must show that A still satisfies property (ii) of Lemma 3.1, with high probability.

Let $n \in [p/5, 7p/5] \cap \mathbb{Z}$. We know that, with high probability, n is the sum of three distinct elements of R in many different ways. Thus, having $n \notin A + A + A$ means that, whenever n is written as $n = a + b + c$ with distinct $a, b, c \in R$, at least one of a, b, c is in X.

Define

$$\mathcal{Q}(n) := \big\{(a, b, c, d, e) \in I^5 : a, b, c \text{ pairwise distinct},\ a + b + c = n,\ c - d - e \in \{0, 1\}\big\}.$$

Suppose that n is such that

$$\big|\mathcal{Q}(n) \cap R^5\big| \geqslant K.$$

We claim that either $\max_{m \in \mathbb{N}} \mathbf{1}_R * \mathbf{1}_R(m) \geqslant 10$, $\max_{m \in \mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) \geqslant 10$, or there are four tuples $(a_i, b_i, c_i, d_i, e_i)_{1 \leqslant i \leqslant 4} \in \mathcal{Q}(n) \cap R^5$ such that, for all $1 \leqslant i < j \leqslant 4$,

$$\{a_i, b_i, c_i, d_i, e_i\} \cap \{a_j, b_j, c_j, d_j, e_j\} = \emptyset. \tag{A.7}$$

Suppose first that there is some $a \in R$ appearing as the first coordinate of 200 tuples

$$(a, b_i, c_i, d_i, e_i)_{1 \leqslant i \leqslant 200} \in \mathcal{Q}(n) \cap R^5.$$

Then $n - a$ can be written as $b_i + c_i$ for all $1 \leqslant i \leqslant 200$. If $\mathbf{1}_R * \mathbf{1}_R(n - a) < 10$, there are $b, c \in R$ such that $b_i = b$ and $c_i = c$ for at least 20 values of $i \in \{1, \ldots, 200\}$, say for $i \in J$ where $|J| \geqslant 20$. In turn, this implies that $d_i + e_i \in \{c - 1, c\}$ for all $i \in J$, but if $\mathbf{1}_R * \mathbf{1}_R(c - 1) < 10$ and $\mathbf{1}_R * \mathbf{1}_R(c) < 10$ there must exist two distinct indices $i, j \in J$ such that $(d_i, e_i) = (d_j, e_j)$. This is impossible since (a, b_i, c_i, d_i, e_i) and (a, b_j, c_j, d_j, e_j) are distinct tuples.

Similar reasoning shows that no $r \in R$ can appear as the second or third coordinate of 200 tuples in $\mathcal{Q}(n) \cap R^5$, unless $\max_{m \in \mathbb{N}} \mathbf{1}_R * \mathbf{1}_R(m) \geqslant 10$.

Suppose that there is some $d \in R$ that appears as the fourth coordinate of 200 tuples

$$(a_i, b_i, c_i, d, e_i)_{1 \leqslant i \leqslant 200} \in \mathcal{Q}(n) \cap R^5.$$

Then $c_i - e_i \in \{d, d + 1\}$ for all i. If $\max_{m \in \mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) < 10$, there are $c, e \in R$ and at least 10 values of i such that $c_i = c$ and $e_i = e$. Thus, for these $\geqslant 10$ values of i, we have $a_i + b_i = n - c$. If $\mathbf{1}_R * \mathbf{1}_R(n - c) < 10$, there are two distinct indices i, j such that $(a_i, b_i, c_i, d, e_i) = (a_j, b_j, c_j, d, e_j)$, a contradiction.

The same reasoning shows that no $e \in R$ appears in 200 distinct elements of $\mathcal{Q}(n) \cap R^5$, unless $\big\|\mathbf{1}_R * \mathbf{1}_R\big\|_\infty \geqslant 10$ or $\max_{m \in \mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) \geqslant 10$.

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

The claim follows easily from these observations. Indeed, suppose that $\|\mathbf{1}_R * \mathbf{1}_R\|_\infty < 10$, $\max_{m\in\mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) < 10$ and $|\mathcal{Q}(n) \cap R^5| \geqslant K$. Let $(a_1, b_1, c_1, d_1, e_1)$ be an arbitrary element of $\mathcal{Q}(n) \cap R^5$. The previous observations show that the number of tuples in $\mathcal{Q}(n) \cap R^5$ having a coordinate in $\{a_1, \dots, e_1\}$ is bounded above by an absolute constant. Since K is assumed to be sufficiently large, we can find another tuple $(a_2, b_2, c_2, d_2, e_2)$ such that $\{a_1, \dots, e_1\} \cap \{a_2, \dots, e_2\} = \emptyset$. Repeating, we find four tuples $(a_i, b_i, c_i, d_i, e_i)_{1\leqslant i\leqslant 4} \in \mathcal{Q}(n) \cap R^5$ satisfying (A.7).

By this claim and the union bound, we deduce that

$$\begin{aligned}
&\mathbb{P}\big(\exists n \in [p/5, 7p/5] \cap \mathbb{Z}, |\mathcal{Q}(n) \cap R^5| \geqslant K\big) \\
&\quad \leqslant \mathbb{P}\big(\|\mathbf{1}_R * \mathbf{1}_R\|_\infty \geqslant 10\big) + \mathbb{P}\big(\max_{m\in\mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) \geqslant 10\big) \\
&\qquad + \sum_{n\in[p/5,7p/5]\cap\mathbb{Z}} \ \sum_{\substack{(a_i,b_i,c_i,d_i,e_i)_{1\leqslant i\leqslant 4}\in\mathcal{Q}(n) \\ \text{satisfying (A.7)}}} \mathbb{P}\big(\forall i,\ a_i, b_i, c_i, d_i, e_i \in R\big).
\end{aligned}$$

By (A.2) and (A.3), the probabilities $\mathbb{P}\big(\|\mathbf{1}_R * \mathbf{1}_R\|_\infty \geqslant 10\big)$ and $\mathbb{P}(\max_{m\in\mathbb{N}} \mathbf{1}_R * \mathbf{1}_{-R}(m) \geqslant 10)$ are $\ll K^{-1}$. By (A.7), the events $\{a_i, \dots, e_i \in R\}$ and $\{a_j, \dots, e_j \in R\}$ are independent for $i \neq j$. Thus, the last probability is

$$\mathbb{P}(\forall i,\ a_i, b_i, c_i, d_i, e_i \in R) = \prod_{i=1}^{4} \mathbb{P}\big(a_i, b_i, c_i, d_i, e_i \in R\big) = \prod_{i=1}^{4} (Kp^{-2/3})^{|\{a_i,\dots,e_i\}|}.$$

For $k \in \{3, 4, 5\}$, let

$$\mathcal{Q}_k(n) = \big\{(a, b, c, d, e) \in \mathcal{Q}(n) : \big|\{a, b, c, d, e\}\big| = k\big\}.$$

It is not hard to see that $|\mathcal{Q}_k(n)| \ll p^{k-2}$ for $k \in \{3, 4, 5\}$. Hence, we obtain

$$\begin{aligned}
\mathbb{P}\big(\exists n \in [p/5, 7p/5] \cap \mathbb{Z},\ |\mathcal{Q}(n) \cap R^5| \geqslant K\big) &\ll K^{-1} + \sum_{n\in[p/5,7p/5]\cap\mathbb{Z}} \prod_{i=1}^{4} \sum_{k=3}^{5} |\mathcal{Q}_k(n)| (Kp^{-2/3})^k \\
&\ll K^{-1} + p\left(\sum_{k=3}^{5} p^{k-2} (Kp^{-2/3})^k\right)^4 \\
&\ll K^{-1} + p(K^5 p^{-1/3})^4 \\
&\ll K^{-1}.
\end{aligned}$$

The last computation and (A.6) imply that there is a set $R \subset I$ such that:

(a) for all $n \in [p/5, 7p/5] \cap \mathbb{Z}$, there are $> K^2$ sets $\{a, b, c\}$ of three distinct elements of R such that $n = a + b + c$;
(b) for all $n \in [p/5, 7p/5] \cap \mathbb{Z}$, there are $\ll K$ sets $\{a, b, c\}$ of three distinct elements of R such that $n = a + b + c$, with one of a, b, c in $R + R + \{0, 1\}$;

since with high probability, both properties hold simultaneously.

As announced earlier, we define $X = R \cap (R + R + \{0, 1\}$ and $A = R \setminus X$. Then A and $A + A + \{0, 1\}$ are disjoint by construction, and for every $n \in [p/5, 7p/5] \cap \mathbb{Z}$ there are $K^2 - O(K) \geqslant 1$ ways to write n as the sum of three elements of R, none of which being in X. This means that we can write n as the sum of three elements of A, and we are done. □

https://doi.org/10.1112/S0010437X24007140 Published online by Cambridge University Press

References

AKS81 M. Ajtai, J. Komlós and E. Szemerédi, *A dense infinite Sidon sequence*, European J. Combin. **2** (1981), 1–11.

Cil14 J. Cilleruelo, *Infinite Sidon sequences*, Adv. Math. **255** (2014), 474–486.

Cil15 J. Cilleruelo, *On Sidon sets and asymptotic bases*, Proc. Lond. Math. Soc. (3) **111** (2015), 1206–1230.

DP09 J.-M. Deshouillers and A. Plagne, *A Sidon basis*, Acta Math. Hungar. **123** (2009).

Erd81 P. Erdős, *Solved and unsolved problems in combinatorics and combinatorial number theory*, European J. Combin. **2** (1981), 1–11.

Erd98 P. Erdős, *The probability method: successes and limitations*, J. Statist. Plann. Inference **72** (1998), 207–213.

ESS94a P. Erdős, A. Sárközy and V. T. Sós, *On additive properties of general sequences*, Discrete Math. **136** (1994), 75–99.

ESS94b P. Erdős, A. Sárközy and V. T. Sós, *On sum sets of Sidon sets, I*, J. Number Theory **47** (1994), 329–347.

ESS95 P. Erdős, A. Sárközy and V. T. Sós, *On sum sets of Sidon sets, II*, Israel J. Math. **90** (1995), 221–233.

Gog75 J. H. Goguel, *Über Summen von zufälligen Folgen natürlicher Zahlen*, J. Reine Angew. Math. **272** (1975), 63–77.

HR83 H. Halberstam and K. F. Roth, *Sequences* (Springer, New York, 1983).

IK04 H. Iwaniec and E. Kowalski, *Analytic number theory*, vol. 53 (American Mathematical Society, 2004).

JRL11 S. Janson, A. Rucinski and T. Luczak, *Random graphs* (John Wiley & Sons, 2011).

Kis10 S. Z. Kiss, *On Sidon sets which are asymptotic bases*, Acta Math. Hungar. **128** (2010), 46–58.

KRS14 S. Z. Kiss, E. Rozgonyi and C. Sándor, *On Sidon sets which are asymptotic bases of order 4*, Funct. Approx. Comment. Math. **51** (2014), 393–413.

KS21 S. Z. Kiss and C. Sándor, *Generalized asymptotic Sidon basis*, Discrete Math. **344** (2021), 112208.

KS22 S. Z. Kiss and C. Sándor, *On $B_h[1]$-sets which are asymptotic bases of order 2h*, Preprint (2022), arXiv:2202.13841.

KS23 S. Z. Kiss and C. Sándor, *Dense sumsets of Sidon sequences*, European J. Combin. **107** (2023), 103600.

Ruz98 I. Z. Ruzsa, *An infinite Sidon sequence*, J. Number Theory **68** (1998), 63–71.

Sár01 A. Sárközy, *Unsolved problems in number theory*, Period. Math. Hungar. **42** (2001), 17–35.

Saw21 W. Sawin, *Square-root cancellation for sums of factorization functions over squarefree progressions in $\mathbb{F}_q[t]$*, Acta Math., to appear. Preprint (2021), arXiv:2102.09730.

Cédric Pilatte cedric.pilatte@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Andrew Wiles Building, Oxford OX2 6GG, UK

Compositio Mathematica is owned by the Foundation Compositio Mathematica and published by the London Mathematical Society in partnership with Cambridge University Press. All surplus income from the publication of *Compositio Mathematica* is returned to mathematics and higher education through the charitable activities of the Foundation, the London Mathematical Society and Cambridge University Press.