INFINITE SIDON SEQUENCES

JAVIER CILLERUELO

Abstract. We present a method to construct dense infinite Sidon sequences based on
the discrete logarithm. We give an explicit construction of an infinite Sidon sequence \( B \)
with \( B(x) = x^{\sqrt{2}-1+o(1)} \). Ruzsa proved the existence of a Sidon sequence with similar
counting function but his proof was not constructive.

Our method generalizes to \( B_h \) sequences: For all \( h \geq 3 \), there is a \( B_h \) sequence \( B \)
such that \( B(x) = x^{(h-1)^2+1-(h-1)+o(1)} \).

1. Introduction

According to Erdős [3], around 1932 Simon Sidon asked him about the growing of
those infinite sequences \( B \) with the property that all sums \( b + b' \), \( b \leq b' \), \( b, b' \in B \) are
distinct. Later Erdős named them Sidon sequences. Sidon had found one with counting
function \( B(x) \gg x^{1/4} \) and Erdős observed that the greedy algorithm, described below,
provides another with \( B(x) \gg x^{1/3} \).

Erdős conjectured that for any \( \epsilon > 0 \) should exist a Sidon sequence with \( B(x) \gg x^{1/2-\epsilon} \), but the sequence given by the greedy algorithm was, for almost 50 years, the
densest example known. That was until that, in 1981, Atjai, Komlos and Szemeredi [1]
proved the existence of an infinite Sidon sequence such that \( B(x) \gg (x \log x)^{1/3} \). The
main tool was a remarkable new result in graph theory that they proved in that seminal
paper. They wrote:

The task of constructing a denser sequence has so far resisted all efforts, both con-
structive and random methods. Here we use a random construction for giving a sequence
which is denser than the above trivial one.

2000 Mathematics Subject Classification. 11B83.

Key words and phrases. Sidon sets, \( B_h \) sequences, Probabilistic method, Discrete logarithm.
This work was supported by Grant MTM 2011-22851 of MICINN (Spain).
So that, it was a surprise when Ruzsa [5] overcome the barrier of the exponent 1/3, proving the existence of an infinite Sidon sequence \( B \) with \( B(x) = x^{\sqrt{2}-1+\theta(1)} \). The starting point of Ruzsa’s approach was the sequence \( \{ \log p \} \) where \( p \) runs over all the prime numbers. Ruzsa’s proof is not constructive. For each \( \alpha \in [1, 2] \) he considered a sequence \( B_\alpha = \{ b_p \} \) where each \( b_p \) is built using the binary digits of \( \alpha \log p \). What Ruzsa proved is that for almost all \( \alpha \in [1, 2] \) the sequence \( B_\alpha \) is nearly a Sidon sequence in the sense that removing not too many elements from the sequence it is possible to destroy all the repeated sums that eventually appear.

Here we present a method to construct explicitly dense infinite Sidon sequences. It is inspired by the finite Sidon set

\[ A = \{ \log_q p : p \text{ prime}, p \leq \sqrt{q} \}, \]

where \( g \) a generator of \( \mathbb{F}_q^* \) and \( \log_q x \) denotes the discrete logarithm of \( x \) in \( \mathbb{F}_q^* \). The set \( A \) is indeed a Sidon set in \( \mathbb{Z}_{q-1} \) with size \( |A| = \pi(\sqrt{q}) \sim 2\sqrt{q}/\log q \). Despite the simplicity of the construction of this finite Sidon set we have not seen it previously in the literature.

Our main result is Theorem 1.2 but to warm up we construct first an infinite Sidon sequence \( B = \{ b_p \}_{p \in \mathbb{P}} \) indexed with all the prime numbers with an easy explicit expression for the elements \( b_p \). This is the first time that an infinite Sidon sequence \( B \) with \( B(x) \gg x^\delta \) for some \( \delta > 1/3 \) is constructed explicitly.

**Theorem 1.1.** Let \( B := B_{q,c} = \{ b_p \}_{p \in \mathbb{P}} \) be the sequence constructed in section 2. We have that for \( c = \frac{3-\sqrt{5}}{2} \) it is an infinite Sidon sequence with

\[ B(x) = x^{\frac{3-\sqrt{5}}{2} + \theta(1)}. \]

Theorem 1.1 is weaker than Theorem 1.2, but we have included it as a separated theorem because the simplicity of the construction. In Theorem 1.2 we construct explicitly a denser infinite Sidon sequence \( B = \{ b_p \}_{p \in \mathbb{P}} \) adding the deletion technique to our method. The starting point is the sequence \( B_{7,c} = \{ b_p \}_{p \in \mathbb{P}} \) with \( c = \sqrt{2} - 1 \). This sequence is not a Sidon sequence but we can delete some elements \( b_p \) from the sequence to destroy the repeated sums that could appear. Thus, the final set of indexes of our sequence will be not the whole set of the prime numbers, as in the construction of Theorem 1.1, but the set \( \mathbb{P}^* \) formed by the survived primes after we remove a thin subsequence of the primes that we can describe explicitly.

**Theorem 1.2.** The sequence \( B = \{ b_p \}_{p \in \mathbb{P}^*} \) constructed in subsection 2.2 is an infinite Sidon sequence with

\[ B(x) = x^{\sqrt{2}-1+\theta(1)}. \]

Note that the exponent of the counting function in the explicit construction of Theorem 1.2 is the same that Ruzsa obtained in his random construction. Furthermore,
it can be checked easily that the algorithm used to construct the Sidon sequence in
Theorem 1.2 is efficient in the sense that only $O(x^{\sqrt{2} - 1 + o(1)})$ elementary operations are
needed to list all the elements $b_p \leq x$.

Our approach also generalizes to $B_h$ sequences, that is, sequences such that all the
sums $b_1 + \cdots + b_h$, $b_1 \leq \cdots \leq b_h$ are distinct. To deal with these cases, however, we
need to introduce a probabilistic argument in an unusual way and it becomes the proof
of the following theorem not constructive.

**Theorem 1.3.** For any $h \geq 3$ there exists an infinite $B_h$ sequence $B$ with

$$B(x) = x^{\sqrt{(h-1)^2 + 1} - (h-1) + o(1)}.$$  

The exponent in Theorem 1.3 is greater than $1/(2h - 1)$, that given by the greedy
algorithm for $B_h$ sequences. It should be mentioned that R. Tesoro and the author
[2] have proved recently Theorem 1.3 in the cases $h = 3$ and $h = 4$ using a variant of
Ruzsa’s approach which makes use of the sequence $\{\theta(p)\}$ of arguments of the Gaussian
primes $p = |p|e^{2\pi i \theta(p)}$ instead of the sequence $\{\log p\}$ considered by Ruzsa. However,
that proof does not extend to all $h$.

In the last section we present an alternative method to construct infinite Sidon se-
quences. It has the same flavor than the construction described in section 2 but the
irreducible polynomials in $F_2[X]$ play the role of the prime numbers in the set of pos-
tive integers. The finite version of this construction is the following. We identify
$F_{2^n} \simeq F_2[X]/q(X)$ where $q := q(X)$ is an irreducible polynomial in $F_2[X]$ with $\deg q = n$.
Let $g$ a generator of $F_{2^n}$ and $g := g(X)$ the corresponding polynomial

$$A = \{ x(p) : g^{x(p)} \equiv p \pmod{q}, \ p \text{ irreducible in } F_2[X], \ \deg(p) < n/2 \}$$

is a Sidon set in $\mathbb{Z}_{2^{n-1}}$ of size $|A| \gg 2^{n/2}/n$.

We present an sketch of how to reprove Theorems 1.1, 1.2 and 1.3 using this alternative
approach.

## 2. THE CONSTRUCTION

Fix $c$, $0 < c < 1/2$ and consider the partition of the set of the prime numbers,

$$\mathcal{P} = \bigcup_{k \geq 2} \mathcal{P}_k, \quad \text{where} \quad \mathcal{P}_k = \left\{ \text{p prime : } \frac{2^{c(k-1)^2}}{k-1} < p \leq \frac{2^{c^2 k^2}}{k} \right\}.$$  

Let $q := q_1 < q_2 < \cdots$ be a given infinite sequence of prime numbers satisfying

$$2^{2j-1} < q_j \leq 2^{2j+1}$$

for all $j \geq 1$ and choose, for each $j$, a primitive root $g_j$ of $F_{q_j}^\ast$. 

Let us define the sequence \( B_{q,c} = \{b_p\}_{p \in \mathcal{P}} \) as follows: for \( p \in \mathcal{P}_k \), set
\[
(2.2) \quad b_p = x_1(p) + \sum_{2 \leq j \leq k} x_j(p)(4q_1) \cdots (4q_{j-1}),
\]
where \( x_j(p), 1 \leq j \leq k \) is the solution of the congruence
\[
g_j^{x_j(p)} = p \pmod{q_j}, \quad q_j + 1 \leq x_j(p) \leq 2q_j - 1.
\]
We define \( x_j(p) = 0 \) when \( j > k \).

2.1. Properties of the sequence \( B_{q,c} \). Let us consider the following well known fact, which will be used in the proofs of Propositions 1 and 2:

Given an infinite sequence of positive integers \( r_1, \ldots, r_j, \ldots \) (the base), any non-negative integer can be written, in an only way, in the form
\[
y_1 + y_2 r_1 + y_3 r_1 r_2 + \cdots + y_j r_1 \cdots r_{j-1} + \cdots
\]
with digits \( 0 \leq y_j < r_j, \ j \geq 1 \).

We observe that if \( p \in \mathcal{P}_k \), the digits of \( b_p \) in the base \( 4q_1, \ldots, 4q_j, \ldots \) are
\[
b_p = x_k(p)x_{k-1}(p) \cdots x_2(p)x_1(p).
\]

**Proposition 1.** All the elements \( b_p \) of the sequence \( B_{q,c} \) are distinct and the counting function satisfies \( B_{q,c}(x) = x^{c+o(1)} \).

**Proof.** Suppose that \( b_p = b_p' \). Thus \( x_j(p) = x_j(p') \) for all \( j \geq 1 \) and we have, by construction, that \( p, p' \in \mathcal{P}_k \) where \( k \) is the largest \( j \) such that \( x_j(p) \neq 0 \). We also know that
\[
p \equiv g_j^{x_j(p)} \equiv g_j^{x_j(p')} \equiv p' \pmod{q_j}
\]
for all \( j \leq k \) and then, \( p \equiv p' \pmod{q_1 \cdots q_k} \). If \( p \neq p' \) we would have
\[
2^{ck^2} \geq |p - p'| \geq q_1 \cdots q_k > 2^{1+3+\cdots+(2k-1)} = 2^{k^2},
\]
which is impossible because \( c < 1 \).

To study the growing of the sequence \( B_{q,c} \), we consider, for any \( x \), the integer \( k \) such that
\[
(4q_1) \cdots (4q_k) < x \leq (4q_1) \cdots (4q_{k+1}).
\]
Using (2.1) we can check that \( 2^{k^2+2k} < x \leq 2^{(k+2)^2+2(k+1)} \) and then \( 2^{k^2} = x^{1+o(1)} \).

We observe that if \( p \leq 2^{ck^2}/k \) then \( p \in \mathcal{P}_j \) for some \( j \leq k \) and then \( b_p \leq (4q_1) \cdots (4q_j) \leq (4q_1) \cdots (4q_k) \leq x \). Thus,
\[
B_{q,c}(x) \geq \pi(2^{ck^2}/k) = 2^{k^2(1+o(1))} = x^{c+o(1)}.
\]

For the upper bound we observe that if \( p > 2^{c(k+1)^2}/(k+1) \) then \( p \in \mathcal{P}_j \) for some \( j \geq k+2 \) and then \( b_p > (4q_1) \cdots (4q_{j-1}) \geq (4q_1) \cdots (4q_{k+1}) \geq x \). Thus
\[
B_{q,c}(x) \leq \pi(2^{c(k+1)^2}/(k+2)) = 2^{ck^2(1+o(1))} = x^{c+o(1)}.
\]
The next proposition concerns to the Sidoness quality of the sequence $\mathcal{B}_{\pi,c}$.

**Proposition 2.** Suppose that there exist $b_{p_1}, b_{p_2}, b_{p_1'}, b_{p_2'} \in \mathcal{B}_{\pi,c}$, $b_{p_1} > b_{p_1'} \geq b_{p_2'} > b_{p_2}$ such that

$$b_{p_1} + b_{p_2} = b_{p_1'} + b_{p_2'}.$$

Then we have that:

i) there exist $k_2, k_1$, $k_2 \leq k_1$ such that $p_1, p_1' \in \mathcal{P}_{k_1}$, $p_2, p_2' \in \mathcal{P}_{k_2}$.

ii) $p_1 p_2 \equiv p_1' p_2' \pmod{q_1 \cdots q_{k_2}}$

iii) $p_1 \equiv p_1' \pmod{q_{k_2+1} \cdots q_{k_1}}$ if $k_2 < k_1$.

iv) $(1 - c) k_1^2 < k_2 < \frac{c}{1 - c} k_1^2$.

**Proof.** Since $0 \leq x_j(p_1) + x_j(p_2) < 4q_j$ for all $j$, the equality $b_{p_1} + b_{p_2} = b_{p_1'} + b_{p_2'}$ implies that the digits of both sums are equal:

$$x_j(p_1) + x_j(p_2) = x_j(p_1') + x_j(p_2')$$

for all $j$. By construction, $p_1 \in \mathcal{P}_{k_1}$ and $p_2 \in \mathcal{P}_{k_2}$ where $k_1$ is the largest $j$ such that

$$x_j(p_1) + x_j(p_2) \geq q_j + 1$$

and $k_2$ is the largest $j$ such that

$$x_j(p_1) + x_j(p_2) \geq 2q_j + 2.$$

This observation proves part i). To prove parts ii) and iii) first we observe that (2.3) implies that for all $j$ we have

$$g_j^{x_j(p_1) + x_j(p_2)} \equiv g_j^{x_j(p_1') + x_j(p_2')} \pmod{q_j}.$$

We also know that if $p \in \mathcal{P}_k$, then $g_j^{x_j(p)} \equiv p \pmod{q_j}$ for $j \leq k$ and $g_j^{x_j(p)} \equiv 1 \pmod{q_j}$ when $j > k$.

Thus, for $j \leq k_2$ we have that $p_1 p_2 \equiv p_1' p_2' \pmod{q_j}$ and then

$$p_1 p_2 \equiv p_1' p_2' \pmod{q_1 \cdots q_{k_2}}.$$

If $k_2 < k_1$, for $k_2 + 1 \leq j \leq k_1$ we have that $p_1 \equiv p_1' \pmod{q_j}$ and then

$$p_1 \equiv p_1' \pmod{q_{k_2+1} \cdots q_{k_1}}.$$

Part ii) and the inequalities on $p_i$ and $q_j$ yield

$$\frac{2^{(k_1^2+k_2^2)}}{k_1 k_2} \geq |p_1 p_2 - p_1' p_2'| \geq q_1 \cdots q_{k_2} > 2^{1+3+\cdots+(2k_2-1)} = 2^{k_2^2} \implies k_2^2 < \frac{c}{1 - c} k_1^2.$$

In particular it implies that $k_2 < k_1$ and we can apply part iii), which gives

$$\frac{2^{k_1^2}}{k_1} \geq |p_1 - p_1'| \geq q_{k_2+1} \cdots q_{k_1} > 2^{(2k_2+1)\cdots+(2k_1-1)} = 2^{k_1^2-k_2^2} \implies k_2^2 > (1 - c)k_1^2.$$
2.2. Proof of Theorems 1.1 and 1.2. To prove the first part of Theorem 1.1 we simply observe that if there is a repeated sum, then Proposition 2, iv) implies that $1 - c < \frac{c}{1 - c}^2$, which does not hold for $c = \frac{3 - \sqrt{5}}{2}$. Thus, $\mathcal{B}_{\eta, c}$ is a Sidon sequence for this value of $c$.

To prove Theorem 1.2 let us consider the sequence $\mathcal{B}_{\eta, c}^* = \{b_p\}_{p \in \mathcal{P}^*}$ with $c = \sqrt{2} - 1$ where the numbers $b_p$ are defined as in (4.2) but where $\mathcal{P}^*$ is not the whole set of the prime numbers but the set of the survived primes after removing a thin subset to avoid the presence of some repeated sums. More precisely, using the notation $Q_1 = q_1 \cdots q_{k_2}$ and $Q_2 = q_{k_2+1} \cdots q_{k_1}$, we define

$$\mathcal{P}^* = \bigcup_{k_1}(\mathcal{P}_{k_1} \setminus \mathcal{R}_{k_1})$$

where the removed set $\mathcal{R}_{k_1}$ consist of all the primes in $\mathcal{P}_{k_1}$ dividing some integer $s \neq 0$ from some of the sets

$$S_{k_2, k_1} = \left\{ s = s_1 Q_1 + s_2 Q_2 : 1 \leq |s_i| \leq \frac{2c(k_2^2 + k_1^2)}{k_2 k_1 Q_i^2}, i = 1, 2 \right\}$$

with $k_2^2 < \frac{c}{1 - c} k_1^2$. That is,

$$\mathcal{R}_{k_1} = \left\{ p \in \mathcal{P}_{k_1} : p | s, \text{ for some } s \in \bigcup_{k_2 < \sqrt{1 - c} k_1} S_{k_2, k_1}, s \neq 0 \right\}.$$ 

We claim that for any $\eta$ and $c = \sqrt{2} - 1$, the sequence $\mathcal{B}_{\eta, c}^* = \{b_p\}_{p \in \mathcal{P}^*}$ is an infinite Sidon sequence with $\mathcal{B}_{\eta, c}^*(x) = x^{\sqrt{2} - 1 + o(1)}$.

To see that $\mathcal{B}_{\eta, c}^*$ is a Sidon sequence, suppose that $b_{p_1} + b_{p_2} = b_{p_1'} + b_{p_2'}$ with $b_{p_1} > b_{p_1'} \geq b_{p_2'} > b_{p_2}$ and $p_1, p_1', p_2, p_2' \in \mathcal{P}^*$. Proposition 2 implies that $p_1, p_1' \in \mathcal{P}_{k_1} \setminus \mathcal{R}_{k_1}$ and $p_2, p_2' \in \mathcal{P}_{k_2} \setminus \mathcal{R}_{k_2}$ for some $k_2, k_1$ satisfying $k_2^2 < \frac{c}{1 - c} k_1^2$. Next, note that thanks to parts ii) and iii) of Proposition 2 we can write

$$p_1(p_2 - p_2') = s_1 Q_1 + s_2 Q_2$$

for the nonzero integers $s_1 = \frac{p_1 p_2 - p_1' p_2'}{Q_1}$ and $s_2 = \frac{(p_1' - p_1) p_2'}{Q_2}$ which satisfy

$$1 \leq |s_i| \leq \frac{2c(k_2^2 + k_1^2)}{k_2 k_1 Q_i^2}, \quad i = 1, 2.$$

It implies that $p_1(p_2 - p_2') \in S_{k_2, k_1}$ for some $k_2$ with $k_2^2 < \frac{c}{1 - c} k_1^2$, so $p_1 \in \mathcal{R}_{k_1}$ and then $p_1 \notin \mathcal{P}_{k_1} \setminus \mathcal{R}_{k_1}$.

To prove that $\mathcal{B}_{\eta, c}^*(x) = x^{\sqrt{2} + o(1)}$ we only need to show that $|\mathcal{R}_{k_1}| = o(|\mathcal{P}_{k_1}|)$. 

□
We claim that for each non zero integer \( s \in S_{k_2, k_1} \) and \( k_1 \) large enough, there exist at most one \( p \in \mathcal{P}_{k_1} \) dividing \( s \). Otherwise, if \( p, p' \mid s \) we would have that
\[
\frac{2^{2c(k_1-1)^2}}{(k_1-1)^2} \leq pp' \leq |s| \leq 2 \cdot \frac{2^{c(k_1^2+k_2^2)}}{k_1 k_2} \leq \frac{2^{2c} k_1^2+1}{k_1^2 \sqrt{c/(1-c)}},
\]
which does not hold for \( k_1 \) large enough since \( 2c > \frac{c}{1-c} \) for \( c < 1/2 \). Therefore, using the estimates
\[
|\mathcal{P}_{k_1}| = \pi \left( \frac{2^{ck_1^2}}{k_1} \right) - \pi \left( \frac{2^{c(k_1-1)^2}/(k_1-1)}{k_1} \right) \gg \frac{2^{ck_1^2}}{k_1^3},
\]
\[
Q_1 Q_2 = q_1 \cdots q_{k_1} > 2^{k_1^2}
\]
and the identity \( \frac{2c}{1-c} - 1 = c \) when \( c = \sqrt{2} - 1 \) we have, for \( k_1 \) large enough, the wanted estimate,
\[
|\mathcal{R}_{k_1}| \leq \sum_{k_2 < \sqrt{\frac{c}{1-c} k_1}} |S_{k_2, k_1}| \leq \sum_{k_2 < \sqrt{\frac{c}{1-c} k_1}} \left( 2 \cdot \frac{2^{c(k_1^2+k_2^2)}}{k_2 k_1 Q_1} \right) \left( 2 \cdot \frac{2^{c(k_1^2+k_2^2)}}{k_2 k_1 Q_2} \right) \leq 4 \sum_{k_2 < \sqrt{\frac{c}{1-c} k_1}} \frac{2^{2c(k_1^2+k_2^2)-k_2^2}}{k_1^2 k_2^2} \ll \frac{2^{2c(k_1^2+k_2^2)-k_2^2}}{k_1^2 k_2} = \frac{2^{ck_1^2}}{k_1^2} \ll \frac{|\mathcal{P}_{k_1}|}{k_1}.
\]

3. Infinite \( B_h \) Sequences

In the following we shall use the same notation with only minor changes. Fix
\[
c = \sqrt{\log(h-1)^2 + 1 - (h-1)}
\]
and let \( \mathcal{P} = \cup_{k \geq 3} \mathcal{P}_k \) where
\[
\mathcal{P}_k = \left\{ p \text{ prime} : 2^{c(k-1)^2} \left( 1-1/\sqrt{\log(k-1)} \right) < p \leq 2^{c(k^2(1-1/\sqrt{\log k})} \right\}.
\]

Let \( \overline{q} := q_1 < q_2 < \cdots \) be a sequence of primes satisfying \( 2^{2j-1} < q_j \leq 2^{2j+1} \) and let \( g_j \) be a generator of \( \mathbb{F}_{q_j}^* \). For \( p \in \mathcal{P}_k \), we define the integer
\[
b_p = x_1(p) + \sum_{2 \leq j \leq k} x_j(p)(h^2 q_1) \cdots (h^2 q_{j-1}),
\]
where \( x_j(p) \) is the solution of the congruence
\[
x_j^{x_j(p)} \equiv p \pmod{q_j}, \quad (h-1)q_j + 1 \leq x_j(p) \leq hq_j - 1.
\]
We define \( x_j(p) = 0 \) for \( j > k \).
We observe that the sequence \( B_{q,c} = \{b_p\} \) will be a \( B_h \) sequence if and only if for any \( l, \ 2 \leq l \leq h \) there not exists a repeated sum in the form

\[
\begin{align*}
&b_{p_1} + \cdots + b_{p_l} = b_{p'_1} + \cdots + b_{p'_l} \\
&\{b_{p_1}, \ldots, b_{p_l}\} \cap \{b_{p'_1}, \ldots, b_{p'_l}\} = \emptyset \\
&b_{p_1} \geq \cdots \geq b_{p_l} \\
&b_{p'_1} \geq \cdots \geq b_{p'_l}.
\end{align*}
\]

The following proposition is just a generalization of Proposition 2.

**Proposition 3.** Suppose that there exist \( p_1, \ldots, p_l, p'_1, \ldots, p'_l \in B_{q,c} \) satisfying (3.1). Then we have:

i) \( p_i, p'_i \in P_{k_i}, \ i = 1, \ldots, l \) for some \( k_1 \leq \cdots \leq k_l \).

ii) \[
\begin{align*}
&\begin{array}{c}
p_1 \cdots p_l \\
p_1 \cdots p_{l-1}
\end{array} \equiv \\
&\begin{array}{c}
p'_1 \cdots p'_l (\text{mod } q_1 \cdots q_{k_l}) \\
p'_1 \cdots p'_{l-1} (\text{mod } q_{k_l+1} \cdots q_{k_{l-1}})
\end{array} \quad \text{if } k_l < k_{l-1} \\
&\cdots \\
&\begin{array}{c}
p_1 \\
p_1
\end{array} \equiv \\
&\begin{array}{c}
p'_1 (\text{mod } q_{k_{l-1}+1} \cdots q_{k_1})
\end{array} \quad \text{if } k_2 < k_1.
\end{align*}
\]

iii) \( k_2 \leq \frac{c}{1-c} (k_1^2 + \cdots + k_{l-1}^2) \).

iv) \( q_1 \cdots q_{k_1} \mid \prod_{i=1}^l (p_1 \cdots p_i - p'_1 \cdots p'_i) \).

**Proof.** The proof is similar to the proof of Proposition 2: Here \( k_i \) is the largest \( j \) such that \( x_j(p_i) + \cdots + x_j(p_l) \geq i((h-1)q_j + 1) \). Part iii) is consequence of the first congruence of part ii). Part iv) is also an obvious consequence of part ii). □

### 3.1. Proof of Theorem 1.3.

The sequence \( B_{q,c} \) defined at the beginning of this section may not be a \( B_h \) sequence. The plan of the proof is to remove from \( B_{q,c} = \{b_p\}_{p \in \mathcal{P}} \) the largest element appearing in each such repeated sum to obtain a true \( B_h \) sequence.

More precisely, we define \( \mathcal{P}^* = \mathcal{P}^*(\bar{q}) \) as the set

\[
\mathcal{P}^* = \bigcup_k (\mathcal{P}_k \setminus \mathcal{R}_k(\bar{q}))
\]

where \( \mathcal{R}_k(\bar{q}) = \{p \in \mathcal{P}_k : b_p \text{ is the largest involved in some equation (3.1)}\} \).

It is then clear that the sequence

\[
B_{q,c}^* = \{b_p\}_{p \in \mathcal{P}^*}
\]

is a \( B_h \) sequence.

We can proceed as in the previous section to deduce that \( B_{q,c}(x) = x^{c+o(1)} \). If in addition, \( |\mathcal{R}_k(\bar{q})| = o(|\mathcal{P}_k|) \), we have that

\[
B_{\bar{q},c}^*(x) \sim B_{q,c}(x) = x^{c+o(1)}.
\]
Thus, the proof of Theorem 1.3 will be completed if we prove that there exists a sequence $\mathbf{q}$ such that $|\mathcal{R}_k(\mathbf{q})| = o(\mathcal{P}_k)$.

For $2 \leq l \leq h$ we write

$$\text{Bad}_l(\mathbf{q}, k_1, \ldots, k_1) = \{(p_1, \ldots, p'_l) : p_i, p'_i \in \mathcal{P}_{k_i}, \ i = 1, \ldots, l \text{ satisfying (3.1)}\}.$$ 

Next let us observe that each $p \in \mathcal{R}_k(\mathbf{q})$ comes from some $(p_1, \ldots, p'_l) \in \text{Bad}_l(\mathbf{q}, k_1, \ldots, k_1)$, $2 \leq l \leq h$, $k_1 \leq \cdots \leq k_1 = k$. Thus,

\begin{equation}
|\mathcal{R}_k(\mathbf{q})| \leq \sum_{l=2}^{h} \sum_{k_l \leq \cdots \leq k_1 = k} |\text{Bad}_l(\mathbf{q}, k_1, \ldots, k_1)|
\end{equation}

It happens that we are not able to give a good upper bound for $|\text{Bad}_l(\mathbf{q}, k_1, \ldots, k_1)|$ for a concrete sequence $\mathbf{q} := q_1 < q_2 < \cdots$, but we can do it in average. If the reader is familiarized with Ruzsa’s work, the sequences $\mathbf{q}$ will play the same role as the real parameter $\alpha$ in Ruzsa’s construction.

We consider the probability space of the sequences $\mathbf{q} := q_1 < q_2 \cdots$ where each $q_j$ is chosen at random uniformly between all the primes in the interval $(2^{2^j-1}, 2^{2^j+1}]$. In particular we use that $\pi(2^{2^k+1}) - \pi(2^{2^k-1}) \gg 2^{2k-1} + O(\log k)$ to deduce that for any primes $q_1 < \cdots < q_{k_1}$ satisfying $2^{2j-1} < q_j \leq 2^{2j+1}$ we have

$$\mathbb{P}(q_1, \ldots, q_{k_1} \in \mathbf{q}) = \prod_{k_1=1}^{k_1} \frac{1}{\pi(2^{2k+1}) - \pi(2^{2k-1})} \leq 2^{-k_1^2 + O(k_1 \log k_1)}.$$

Thus, for a given $(p_1, \ldots, p'_l)$, we use Proposition 3, iv) and the estimate $\tau(n) = n^{O(1/\log \log n)}$ for the divisor function to deduce that

$$\mathbb{P}((p_1, \ldots, p'_l) \in \text{Bad}_l(\mathbf{q}, k_1, \ldots, k_1)) \leq \sum_{q_1, \ldots, q_{k_1}} \sum_{q_1 \cdots q_{k_1} \mid p_1 - p_1' \cdots p'_l} \mathbb{P}(q_1, \ldots, q_{k_1} \in \mathbf{q})$$

$$\leq \tau \left( \prod_{i=1}^{l} (p_i - p'_i) \right) 2^{-k_1^2 + O(k_1 \log k_1)} \leq 2^{-k_1^2 + O(k_1^2/\log k_1)}.$$
Thus, using Proposition 3 iii) in the last inequality we have:
\[
\mathbb{E}(\{(p_1, \ldots, p_t) : p_i, p'_i \in \mathcal{P}_{k_i}, \ i = 1, \ldots, t \text{ satisfying (3.1)}\}) \\
\leq 2^{-k_1^2 + O(k_1^2/\log k_1)} \#\{(p_1, \ldots, p_t) : p_i, p'_i \in \mathcal{P}_{k_i}\} \\
\leq 2^{-k_1^2 + O(k_1^2/\log k_1)} |\mathcal{P}_{k_1}|^2 \cdots |\mathcal{P}_{k_1}|^2 \\
\leq 2^{-k_1^2 + O(k_1^2/\log k_1)} \cdot 2^{2c(k_1^2 + \cdots + k_1^2)} - 2k_1^2/\sqrt{\log k_1} \\
\leq 2^{-k_1^2 + \frac{2c}{1-e}(k_1^2 + \cdots + k_1^2)} - (2c + o(1))k_1^2/\sqrt{\log k_1} \\
\leq 2^{-1 + \frac{2c}{1-e}(k_1^2 + \cdots + k_1^2)} - (2c + o(1))k_1^2/\sqrt{\log k_1}.
\]

Using (3.2) we have
\[
\mathbb{E}(|\mathcal{R}_k(\overline{q})|) \leq 2^{-1 + \frac{2c(h-1)}{1-e} - c} = 0 \text{ for } c = \sqrt{(h-1)^2 + 1} - (h-1)
\]
Finally we use that \(-1 + \frac{2c(h-1)}{1-e} - c = 0\) for \(c = \sqrt{(h-1)^2 + 1} - (h-1)\) to obtain
\[
\mathbb{E}\left(\sum_k \frac{|\mathcal{R}_k(\overline{q})|}{|\mathcal{P}_k|}\right) \leq \sum_k k^2 2^{-1 + \frac{2c(h-1)}{1-e} - c} k^2 - (c + o(1))k^2/\sqrt{\log k} \\
\leq \sum_k k^2 2^{-(c + o(1))k^2/\sqrt{\log k}}.
\]

Since the series is convergent we have that for almost all sequences \(\overline{q}\) the series
\[
\sum_k \frac{|\mathcal{R}_k(\overline{q})|}{|\mathcal{P}_k|}
\]
is convergent. Therefore, for any of these \(\overline{q}\) we have that \(|\mathcal{R}_k(\overline{q})| = o(|\mathcal{P}_k|)\), which is what we wanted to prove.

4. An alternative construction

The theorems proved in this paper could have been proved using the following alternative construction, which, although has the same flavor than the one described in section 2, it uses the irreducible polynomials in \(\mathbb{F}_2[X]\) instead of the prime numbers.

Let \(\overline{q} = \{q_j\}\) be any infinite sequence of irreducibles polynomials in \(\mathbb{F}_2[X]\) of degree \(\deg(q_j) = 2j - 1\). For each \(j\), let \(g_j(X)\) a generator of \(\mathbb{F}_2[X]/q_j(X)\). Fix \(c, \ 0 < c < 1/2\) and for each \(k \geq 2\), let
\[
P_k = \{p \text{ irreducible polynomials in } \mathbb{F}_2[X] : c(k-1)^2 < \deg p \leq ck^2\}.
\]

Consider the sequence \(\mathcal{B}_{\overline{q},c} = \{b_p\}\) where, for any \(p(X) \in \mathcal{P}_k\) (we write \(p := p(X)\) for short), the element \(b_p\) is defined by
\[
b_p = \sum_{1 \leq j \leq k} x_j(p)2^{j^2-1}
\]
and $x_j(p)$ is the solution of the polynomial congruence

$$g_j(X)^{x_j(p)} \equiv p(X) \pmod{q_j(X)}, \quad 2^{2j-1} + 1 \leq x_j(p) \leq 2^{2j} - 1.$$ 

Let us define $x_j(p) = 0$ for $j > k$. Note that

$$b_p = x_1(p) + \sum_{2 \leq j \leq k} x_j(p)(4 \cdot 2^1) \cdots (4 \cdot 2^{j-3})$$

and then the digits of $b_p$ in the base $4 \cdot 2^1, \ldots, 4 \cdot 2^{2j-1}, \ldots$ are $b_p = x_k(p) \cdots x_1(p)$.

We use that the number of irreducible polynomials of degree $j$ in $\mathbb{F}_2[X]$ is $\gg 2^{j}/j$ to deduce easily that in this case we also have $B_{q,c}(x) = x^{\alpha(1)}$. Proposition 2 also works here, except that now the congruences are in $\mathbb{F}_2[X]$. It is then very easy to adapt the proofs of Theorems 1.1 and 1.2 to this new construction.

The proof of Theorem 1.3 using this construction is also similar to that given in section 2, except that now we define

$$b_p = x_1(p) + \sum_{2 \leq j \leq k} x_j(p)(h^2 \cdot 2^1) \cdots (h^2 \cdot 2^{2j-3}),$$

where $x_j(p)$ is the solution of the congruence

$$g_j^{x_j(p)} \equiv p(X) \pmod{q_j(X)}, \quad (h - 1)2^{2j-1} + 1 \leq x_j(p) \leq h2^{2j-1} - 1.$$ 

Perhaps, the less known ingredient needed in the proof may be the upper bound $\tau(r(X)) \leq 2^{O(n/\log n)}$ for the number of the divisors of a polynomial $r(X) \in \mathbb{F}_2[X]$ of degree $n$ (see [4]).

**References**


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

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