ESSENTIAL COMPONENTS IN VECTOR SPACES OVER FINITE FIELDS

ZHENCHAO GE AND THÁI HOÀNG LÊ

ABSTRACT. A subset H of non-negative integers is called an essential component, if $\underline{d}(A+H) > \underline{d}(A)$ for all $A \subset \mathbb{N}$ with $0 < \underline{d}(A) < 1$, where $\underline{d}(A)$ is the lower asymptotic density of A. How sparse can an essential component be? This problem was solved completely by Ruzsa. Here, we generalize the problem to the additive group $(\mathbb{F}_p[t], +)$, where p is prime. Our result is analogous to but more precise than Ruzsa's result in the integers. Like Ruzsa's, our method is probabilistic. We also construct an explicit example of an essential component in $\mathbb{F}_p[t]$ with small counting function, based on a construction of small-bias sample space by Alon, Goldreich, Håstad, and Peralta.

1. INTRODUCTION

1.1. **Essential components in $\mathbb{N}$.** Let $\mathbb{N}$ denote the set of nonnegative integers. If $A \subset \mathbb{N}$, the lower asymptotic density of A is defined as

$$\underline{d}(A) = \liminf_{N\to\infty} \frac{|A \cap \{1, 2, \dots, N\}|}{N}$$

and the Shnirelmann density of A is

$$\sigma(A) = \inf_{N\geq 1} \frac{|A \cap \{1, 2, \dots, N\}|}{N}.$$

For two subsets A, B of an abelian group, we define the sumset $A + B = \{a + b : a \in A, b \in B\}$. If $n \in \mathbb{N}$, then nA denotes the n-fold sumset of A. If $A \subset \mathbb{N}$, let $A(x) := \#\{1 \leq n \leq x : n \in A\}$ denote the counting function of A.

A set $H \subset \mathbb{N}$ is called an *essential component* if for any $A \subset \mathbb{N}$ with $0 < \underline{d}(A) < 1$, we have $\underline{d}(A) < \underline{d}(A + H)$. The notion of essential components was introduced by Khinchin [8], though instead of $\underline{d}$ he used σ. For a detailed account of essential components, see [7, Chapter I, §5]. As was proved by Plünnecke [12, Theorem 77, p. 116], H is an essential component with respect to σ if and only if H is an essential component with respect to $\underline{d}$ and $\{0, 1\} \subset H$.

Shnirelmann's inequality [13, Theorem 4.2.1] implies that if $\sigma(H) > 0$ and $0 \in H$, then H is an essential component. Khinchin [8] proved that the set $\{n^2 : n \in \mathbb{N}\}$ is an essential component and Erdős [4] proved that if H is an additive basis of $\mathbb{N}$, i.e. $kH = \mathbb{N}$ for some $k \in \mathbb{Z}$, then H is an essential component. If $kH = \mathbb{N}$ then clearly $H(x) \gg x^{1/k}$. It is natural to ask if H is an essential component, then how small can $H(x)$ be. Linnik [11] constructed an example of an essential component H such that $H(x) = O(\exp(\log^{\frac{9}{10}} x))$. For any given $\eta > 0$, Wirsing [18] constructed an essential component H such that

$$H(x) = O\left(\exp\left(\eta\sqrt{\log x}\log\log x\right)\right). \tag{1}$$

Finally, Ruzsa [14] gave a complete answer to this question by proving the following theorems.

Theorem 1. *For any $c > 0$, there exists an essential component H such that $H(x) \ll \log^{1+c} x$.*

Theorem 2. *Suppose $H \subset \mathbb{N}$ is such that for any $\epsilon > 0$, $|H(x)| < \log^{1+\epsilon} x$ infinitely often. Then there is a set $A \subset \mathbb{N}$ such that*

$$0 < \underline{d}(A) = \underline{d}(A + H) < 1. \tag{2}$$

Consequently, there does not exist an essential component H such that $H(x) \ll \log^{1+o(1)} x$.

The construction in Theorem 1 is probabilistic and no deterministic construction of H is known. Wirisng's bound (1) remains the best explicit construction to date.

1.2. **Essential component in vector spaces.** In view of the influential finite field model in additive combinatorics, it is natural to study the analog of essential components when $\mathbb{N}$ is replaced by a vector space over a finite field.

Let $\mathbb{F} = \mathbb{F}_p$ be the finite field over p elements, where p is prime. Let

$$G := \bigoplus_{i=0}^{\infty} \mathbb{F} = \{(x_0, x_1, \ldots) : x_i \in \mathbb{F}, x_i \neq 0 \text{ for finitely many i}\}.$$

Additively, G is isomorphic to the group $\mathbb{F}[t]$ of polynomials over $\mathbb{F}$. We will write $\mathbb{F}[t]$ and G interchangeably and refer to elements of G as both vectors and polynomials, though no arithmetic structure of $\mathbb{F}[t]$ is involved. Let $G_n = \{x \in \mathbb{F}[t] : \deg x < n\}$, then as an additive group, $G_n \cong \mathbb{F}^n$. We also define $G_0 = \{0\}$. If A is a subset of G, then by A_n we denote $A \cap G_n$. We define the lower asymptotic density of A to be

$$\underline{d}(A) = \liminf_{n\to\infty} \frac{|A_n|}{p^n}.$$

The upper asymptotic density $\overline{d}$ and asymptotic density d are defined similarly. We say a set $H \subset G$ is an essential component if whenever $0 < \underline{d}(A) < 1$, we have

$$\underline{d}(A) < \liminf_{n\to\infty} \frac{|A_n + H_n|}{p^n}.$$

Note that $\liminf_{n\to\infty} \frac{|A_n+H_n|}{p^n}$ is not necessarily the same as $\underline{d}(A + H) = \liminf_{n\to\infty} \frac{|(A+H)_n|}{p^n}$. In contrast to $\mathbb{N}$, G is a group and in general we have $A_n + H_n \subsetneq (A + H)_n$. Since A and H are both infinite sets, little else can be said about $(A + H)_n$ in terms of A_n and H_n. This observation, made precise by the following Proposition, shows that $\underline{d}(A+H)$ is of little interest and our notion is a natural analog of the notion of essential components in $\mathbb{N}$.

Proposition 3. *If $H \subset G$ is infinite, then there is a set $A \subset G$ such that $d(A) = 0$ and $A + H = G$.*

Proof. Since H is infinite, we can find a sequence $(h_n)_{n=1}^{\infty} \subset H$ such that $\deg(h_n) > \max(\deg(h_{n-1}), 2n)$ for any $n > 1$. Let

$$A := \cup_{n=1}^{\infty}(G_n - h_n).$$

Then for any n, $A + H \supset (G_n - h_n) + h_n = G_n$, showing that $A + H = G$. On the other hand, notice that every element in $G_n - h_n$ has degree equal to $\deg(h_n)$. Thus

$$\overline{d}(A) = \lim_{n\to\infty} \frac{|\cup_{j=1}^{n} (G_j - h_j)|}{p^{\deg(h_n)}} = \lim_{n\to\infty} \frac{\sum_{j=1}^{n} p^j}{p^{\deg(h_n)}} = 0.$$

□

The problem of essential components in $\mathbb{F}[t]$ was already studied by Burke [3], who proved the following analog of Erdős' theorem: If H is a basis of order $\leq k$, that is, $kH_n = G_n$ for any $n \in \mathbb{Z}^+$, then H is an essential component. Clearly, if H is a basis of order $\leq k$ then $|H_n| \gg p^{n/k}$.

In this paper, we prove the following analogs of Theorems 1 and 2.

Theorem 4. *For any $c > 0$, there exists an essential component $H \subset G$ such that $|H_n| \ll n^{1+c}$.*

Theorem 5. *Suppose $H \subset G$ is such that for any $\epsilon > 0$, $|H_n| < n^{1+\epsilon}$ infinitely often. Then for any $0 < \delta < 1$, there is a set $A \subset G$ such that*

$$\delta = \underline{d}(A) = \liminf_{n\to\infty} \frac{|A_n + H_n|}{p^n}. \tag{3}$$

Consequently, there does not exist an essential component H such that $|H_n| \ll n^{1+o(1)}$.

We remark that our conclusion (3) is more precise than Ruzsa's (2) in that the density of A can be any prescribed number δ. The proofs of Theorems 4 and 5 will parallel those of Theorems 1 and 2. In our proofs many details are cleaner thanks to the vector space structure of G_n, but some of the arguments don't carry to G_n in a straightforward way, not least because of the fact that there is no linear ordering on G. In proving Theorem 5, we adapt Ruzsa's idea of "niveau sets", namely the set of points at which the Fourier transform of a function is large. The idea was first introduced by Ruzsa in proving Theorem 2 and has found applications in other problems (see [15], [9], [19]) and in particular in vector spaces ([19]). In the context of vector spaces, niveau sets are particularly pleasant.

Similarly to Theorem 1, the construction in Theorem 4 is probabilistic. It is therefore desirable to have an explicit example of an essential component with small counting function. It turns out that there is a connection between essential components in $\mathbb{F}[t]$ and *small-bias sample spaces*, an important notion in theoretical computer science. Using a construction of small-bias sample space by Alon-Goldreich-Håstad-Peralta [1], we prove the following:

Theorem 6. *There exists an essential component $H \subset G$ with counting function $|H_n| = O_p(n^3)$.*

Note that this bound is better than the bound (1) given by Wirsing's construction in $\mathbb{N}$.

The organization of the paper is as follows. In Section 2 we will recall some tools that are used in the proofs. Theorems 4, 5 are proved in Sections 3, 4 respectively. In Section 5 we will discuss explicit constructions of essential components in $\mathbb{F}_p[t]$ and prove Theorem 6.

Acknowledgements. The second author is supported by National Science Foundation Grant DMS-1702296.

2. Preliminaries

2.1. **Notation.** Recall that we use G and $\mathbb{F}[t]$ interchangeably and an element of G can be viewed as both a vector and a polynomial. An element $x = (x_0, x_1, \ldots)$ of G is identified with the polynomial $\sum_{i=0}^{\infty} x_i t^i$. In particular, by $\deg x$, we mean the largest n such that $x_n \neq 0$. We define the *support* of x as $\mathrm{supp}(x) = \{i : x_i \neq 0\}$. We say that x is supported on a set I if $\mathrm{supp}(x) \subset I$. We define $e(x) = e^{2\pi i x}$ for $x \in \mathbb{R}$ and $e_p(x) = e(x/p)$ for $x \in \mathbb{F}$ (so e_p is an additive character on $\mathbb{F}$). We will often make use

of the following fact (where $\cdot$ denotes the scalar product):

$$\sum_{f\in G_n} e_p(x\cdot f) = \begin{cases} p^n, & \text{if } \mathrm{supp}(x)\cap[0,n)=\emptyset \\ 0, & \text{otherwise.}\end{cases} \tag{4}$$

2.2. **Probability tools.**

Lemma 7 (Berry-Esseen inequality [20, Chapter 7, Theorem 6.1]). *Let $X, \{X_j\}_{j=1}^n$ be independent, identically distributed random variables. Let*

$$F(x) = \mathbf{P}\left(\frac{\sum_{j=1}^n X_j - n\mathbf{E}(X)}{\sqrt{n\mathbf{Var}(X)}} \le x\right), \tag{5}$$

and let $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-t^2/2}dt$ be the cumulative distribution function of the standard normal distribution. Suppose $\mathbf{E}(|X-\mathbf{E}(X)|^3) \le K < \infty$. Then

$$\sup_x |F(x)-\Phi(x)| \le \frac{C\cdot K}{n^{1/2}\mathbf{Var}(X)^{3/2}} \tag{6}$$

where C is a constant less than 0.8.

Our next tool is Bernstein's inequality. For real random variables, this can be found in [2, Corollary 2.11]. The complex case follows easily from applying the real case to the real and imaginary parts of Z_j.

Lemma 8 (Bernstein's inequality). *Let $\{Z_j\}_{j=1}^n$ be independent bounded complex random variables such that $\mathbf{E}(\sum Z_j) = A$ and $|Z_j - \mathbf{E}(Z_j)| \le k$ for all $j = 1,\cdots,n$. Suppose $\sum_{j=1}^n \mathbf{Var}(Z_j) \le \sigma^2$. Then for all $\lambda > 0$,*

$$\mathbf{P}\left(|\sum_{j=1}^n Z_j - A| \ge \lambda\right) \le 4\exp\left(\frac{-\lambda^2}{4(\sigma^2+k\lambda/3)}\right).$$

We also need the following version of the law of large numbers.

Lemma 9 (Kolmogorov's strong law of large numbers [21, p. 12]). *Let $\{X_n\}$ be a sequence of independent random variables with $\mathbf{E}(X_n) = 0$ for all n. Let $\{a_n\}$ be a non-decreasing unbounded sequence of positive numbers. If $\sum_{n=1}^\infty \frac{\mathbf{E}(|X_n|^2)}{a_n^2} < \infty$, then*

$$\lim_{n\to\infty} \frac{\sum_{j=1}^n X_j}{a_n} = 0 \quad \text{a. s.} \tag{7}$$

2.3. **Fourier analysis tools.** We need the following lemma of Ruzsa which relates essential components to the Fourier transform. Ruzsa proved it for general abelian groups, though we only need it for the case of G_n.

Lemma 10 ([14, Corollary 7.3]). *Let $K \subset G_n$ and arbitrary complex numbers $(c_k)_{k\in K}$ such that $\sum_{k\in K} c_k = 1$. Define*

$$\xi(x) = \sum_{k\in K} c_k e_p(k\cdot x)$$

for any $x \in G_n$. Suppose there is $\eta > 0$ such that

$$|\xi(x)| \le \eta \text{ for all } x \in G_n, x \ne 0. \tag{8}$$

Then for any set $A \subset G_n$, we have

$$|A+K| \geq |A| + (1-\eta^2)|A|\left(1 - \frac{|A|}{p^n}\right).$$

Remark 1. Lemma 10 says that, if there is a trigonometric polynomial supported on K, all of whose values (except the one at 0) are small, then K serves as an essential component in G_n. The most obvious choice for (c_k) is $c_k = \frac{1}{|K|}$; however, in our application we will have to choose a different function.

Remark 2. Let $S \subset G_n$ be a multiset whose underlying set is K. For $k \in K$, let $c_k = (\text{multiplicity of } k \text{ in } S)/|S|$. It is easy to see that the condition (8) is satisfied if we have

$$\forall c \in \mathbb{F}, \forall x \in G_n, x \neq 0, \qquad \left|\frac{1}{|S|}\#\{s \in S : x \cdot s = c\} - \frac{1}{p}\right| \leq \epsilon \tag{9}$$

with $\epsilon = \frac{\eta}{p}$. A multiset S satisfying (9) is called an *ϵ-biased sample space*, or an *ϵ-biased sample set* (see e.g. [6]). (In the usual definition in the literature, one has $p = 2$, but clearly (9) makes sense for any p.) Thus if a multiset $S \subset G_n$ is an ϵ-biased, then its underlying set K is an essential component in G_n.

For completeness we reproduce Ruzsa's proof of Lemma 10 here.

Proof of Lemma 10. Let $B := (A+K)^c$, then $B \cap (A+K) = \emptyset$. Therefore,

$$\begin{aligned} 0 &= \sum_{x \in G_n} \sum_{b \in B} e_p(-x \cdot b) \sum_{a \in A} e_p(x \cdot a) \sum_{k \in K} c_k e_p(k \cdot x) \\ &= \sum_{x \in G_n} \xi(x) \sum_{b \in B} e_p(-x \cdot b) \sum_{a \in A} e_p(x \cdot a). \end{aligned}$$

By separating the contribution of $x = 0$, we have

$$\begin{aligned} |B||A| &= -\sum_{\substack{x \in G_n \\ x \neq 0}} \xi(x) \sum_{b \in B} e_p(-x \cdot b) \sum_{a \in A} e_p(x \cdot a) \\ &\leq \eta \sum_{\substack{x \in G_n \\ x \neq 0}} \left| \sum_{b \in B} e_p(-x \cdot b) \sum_{a \in A} e_p(x \cdot a) \right| \\ &\leq \eta \left(\sum_{\substack{x \in G_n \\ x \neq 0}} \left| \sum_{b \in B} e_p(-x \cdot b) \right|^2 \right)^{1/2} \left(\sum_{\substack{x \in G_n \\ x \neq 0}} \left| \sum_{a \in A} e_p(x \cdot a) \right|^2 \right)^{1/2} \qquad \text{by Cauchy-Schwarz's inequality} \\ &= \eta \left(|A|(p^n - |A|)\right)^{1/2} \left(|B|(p^n - |B|)\right)^{1/2} \qquad \text{by Plancherel's identity} \end{aligned}$$

Therefore, $|A||B| \leq \eta^2(p^n - |A|)(p^n - |B|)$ and

$$|B| \leq \frac{\eta^2 p^n (p^n - |A|)}{|A| + \eta^2(p^n - |A|)} = p^n \frac{\eta^2(1-\delta)}{\delta + \eta^2(1-\delta)}$$

where $\delta := \frac{|A|}{p^n}$. Since $|B| = p^n - |A+K|$, we have

$$\begin{aligned}\frac{|A+K|}{p^n} &\geq \frac{\delta}{\delta + \eta^2(1-\delta)} \\ &= \delta + \frac{(1-\eta^2)\delta(1-\delta)}{\delta + \eta^2(1-\delta)} \\ &\geq \delta + (1-\eta^2)\delta(1-\delta),\end{aligned}$$

where we applied $\delta + \eta^2(1-\delta) \leq \delta + (1-\delta) = 1$. □

2.4. **Combinatorics tools.**

Lemma 11. *Let $n \in \mathbb{Z}^+$ and $C \subset G_n$ be a subset of G_n with $|C| = \delta p^n > 0$. Then exists $x \in G_n$ such that*

$$|(C-x) \cap G_m| \geq \delta p^m \tag{10}$$

for all $0 \leq m \leq n$. In particular, $x \in C$.

The proof of this lemma can be found in [10, p. 12]. For completeness we include the proof here.

Proof. We prove the lemma by induction on n. When $n = 1$ we can take x to be any element of C. Suppose the lemma is true for subsets of G_{n-1}. Since we have the partition

$$G_n = \cup_{\alpha \in \mathbb{F}} (G_{n-1} + \alpha t^{n-1}),$$

there must be $\alpha \in \mathbb{F}$ such that $|C \cap (G_{n-1} + \alpha t^{n-1})| \geq \delta p^{n-1}$. Therefore, $|(C - \alpha t^{n-1}) \cap G_{n-1}| \geq \delta p^{n-1}$. Applying the induction hypothesis to the set $(C - \alpha t^{n-1}) \cap G_{n-1}$, we see that there is $y \in G_{n-1}$ such that

$$|(C - \alpha t^{n-1} - y) \cap G_m| \geq \delta p^m \tag{11}$$

for all $0 \leq m \leq n-1$. Therefore, (10) is true with $x = \alpha t^{n-1} + y$. The assertion $x \in C$ follows from applying (10) with $m = 0$. □

3. Proof of Theorem 4

In this section, we fix $0 < c < 1$. Let $(X_f)_{f \in G}$ be a family of independent random variables taking values in $\{0, 1\}$ and

$$b_f = \mathbf{P}(X_f = 1) = \frac{\deg(f)^c}{p^{\deg(f)}} \tag{12}$$

if $\deg(f) \geq 1$; $b_f = 1$ if $\deg(f) \leq 0$. Then the X_f's are Bernoulli and

$$\mathbf{E}(X_f) = b_f, \qquad \mathbf{Var}(X_f) = b_f(1-b_f). \tag{13}$$

Now we define

$$H := \{f \in G : X_f = 1\}. \tag{14}$$

On the one hand, we claim that $|H_n| \ll n^{1+c}$ holds almost surely. In order to see this, we apply Lemma 9 to the independent random variables $Y_n = \sum_{\deg(f)=n} X_f - n^c(1-p^{-1})$ and the sequence $a_n = n^{1+c}$ for $n \geq 1$. Since $\mathbf{E}(Y_n) = 0$ for all $n \geq 1$ and $\sum_{n=1}^{\infty} a_n^{-2} \mathbf{E}(|Y_n|^2) \leq \sum_{n=1}^{\infty} n^{-2-c} < \infty$, Lemma 9 implies that

$$\lim_{n\to\infty} \frac{|H_{n+1}| - \mathbf{E}(|H_{n+1}|)}{n^{1+c}} = \lim_{n\to\infty} \frac{\sum_{j=1}^{n} Y_j}{n^{1+c}} = 0 \qquad \text{a.s.} \tag{15}$$

Thus, as $n \to \infty$, $||H_n| - \mathbf{E}(|H_n|)| = o(n^{1+c})$ and $|H_n| \ll \mathbf{E}(|H_n|) \ll n^{1+c}$ holds almost surely.

On the other hand, we will prove that H is an essential component of G almost surely. This is the purpose of the remaining of this section.

The strategy is to use Lemma 10 and produce a trigonometric polynomial supported on H_n, all of whose values are small except the one at 0. A first step is the following, which guarantees that the trigonometric polynomial is small on a set S, as long as $|S|$ is not too big.

Lemma 12. *Let $0 < c < 1$ and n be sufficiently large depending on c. For $f \in G_n$, define*

$$w_0(f) = \frac{1}{p^n b_f} \tag{16}$$

(recall that $b_f = \mathbf{E}(X_f)$). Let

$$\xi_0(x) = \sum_{f \in G_n} w_0(f) X_f e_p(f \cdot x) \tag{17}$$

for $x \in G_n$. Then for any subset $S \subset G_n \setminus \{0\}$ with $|S| \le \exp(\frac{n^c}{200})$, we have

$$\mathbf{P}\left(\left\{|\xi_0(0) - 1| < \frac{1}{3}\right\} \wedge \left\{\max_{x \in S} |\xi_0(x)| < \frac{1}{3}\right\}\right) \ge 1 - \exp\left(\frac{-n^c}{400}\right). \tag{18}$$

Proof of Lemma 12. By the definition of $w_0(f)$, for every $x \in G_n$, we have

$$\mathbf{E}(\xi_0(x)) = \frac{1}{p^n} \sum_{f \in G_n} e_p(x \cdot f) = \begin{cases} 0, & \text{if } x \neq 0 \\ 1, & \text{if } x = 0. \end{cases} \tag{19}$$

For every $x \in G_n$, we have

$$\begin{aligned} \mathbf{Var}(\xi_0(x)) &= \mathbf{Var}(\mathrm{Re}(\xi_0(x))) + \mathbf{Var}(\mathrm{Im}(\xi_0(x))) \\ &\le 2 \sum_{f \in G_n} w_0(f)^2 b_f (1 - b_f) \le \frac{2}{p^{2n}} \sum_{f \in G_n \setminus G_0} \frac{1}{b_f} < \frac{2}{p^{2n}} \sum_{j=1}^{n-1} \frac{p^{2j}}{j^c}. \end{aligned} \tag{20}$$

Note that since $\frac{p^{2(j+1)}}{(j+1)^c} / \frac{p^{2j}}{j^c} \ge \frac{p^2}{2} \ge 2$, it is easy to show that $\sum_{j=1}^{n-1} \frac{p^{2j}}{j^c} \le \frac{p^{2n}}{n^c}$. Hence, for every $x \in G_n$ the variance is

$$\mathbf{Var}(\xi_0(x)) < 2n^{-c}. \tag{21}$$

Moreover, since $|w_0(f) e_p(f \cdot x)(X_f - \mathbf{E}(X_f))| \le 2w_0(f) \le 2n^{-c}$, Bernstein's inequality (Lemma 8) implies that

$$\begin{aligned} &\mathbf{P}\left(|\xi_0(x)| \ge \tfrac{1}{3}\right) \le 4 \exp\left(\tfrac{-n^c}{80}\right) \qquad \text{for } x \neq 0, \\ &\mathbf{P}\left(|\xi_0(0) - 1| \ge \tfrac{1}{3}\right) \le 4 \exp\left(\tfrac{-n^c}{80}\right). \end{aligned} \tag{22}$$

Since $4(|S|+1) \le 4\exp(\frac{n^c}{200}) + 4 < \exp(\frac{n^c}{100})$ holds for all sufficiently large n depending on c, using (22) and the union bound, we obtain that

$$\mathbf{P}\left(\left\{\max_{x \in S} |\xi_0(x)| \ge \frac{1}{3}\right\} \vee \left\{|\xi_0(0) - 1| \ge \frac{1}{3}\right\}\right) \le 4(|S|+1) \exp\left(\frac{-n^c}{80}\right) < \exp\left(\frac{n^c}{100}\right) \exp\left(\frac{-n^c}{80}\right).$$

In other words,

$$\mathbf{P}\left(\left\{\max_{x \in S} |\xi_0(x)| < \frac{1}{3}\right\} \wedge \left\{|\xi_0(0) - 1| < \frac{1}{3}\right\}\right) \ge 1 - \exp\left(\frac{n^c}{100}\right) \exp\left(\frac{-n^c}{80}\right) = 1 - \exp\left(\frac{-n^c}{400}\right),$$

as desired. □

The trigonometric polynomial ξ_0 given by Lemma 12 covers only a set S whose size is small compared to G_n. In the next Lemma, we will produce different trigonometric polynomials ξ_u, each covering a different set S_u, then "glue" these trigonometric polynomials together. We can do this as long as $|S_u|$ is not too big, and no element of S_u is supported on $[n-u, n)$.

Lemma 13. *Let $0 < c < 1$ and n be sufficiently large depending on c. Let u be an integer with $1 \le u < n^{1-c/3}$. For $f \in G_n$ we define*

$$w_u(f) = \begin{cases} w = ((1-p^{-1})\sum_{j=n-u}^{n-1} j^c)^{-1}, & \text{if } n-u \le \deg(f) < n, \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Further, for $x \in G_n$, we define

$$\xi_u(x) = \sum_{f \in G_n} w_u(f) X_f e_p(f \cdot x). \tag{24}$$

Then for any subset $S_u \subset \{x \in G_n : \mathrm{supp}(x) \cap [0, n-u) \neq \emptyset\}$ with $|S_u| \le \exp(\frac{un^c}{2000})$, we have

$$\mathbf{P}\left(\left\{|\xi_u(0) - 1| < \frac{1}{3}\right\} \wedge \left\{\max_{x \in S_u} |\xi_u(x)| < \frac{1}{3}\right\}\right) \ge 1 - \exp\left(\frac{-un^c}{6000}\right). \tag{25}$$

Proof of Lemma 13. We first see that $\mathbf{E}(\xi_u(0)) = \sum_{f \in G_n} w_u(f) b_f = \sum_{j=n-u}^{n-1} \sum_{\deg(f)=j} w b_f = 1$.

For $\mathrm{supp}(x) \cap [0, n-u) \neq \emptyset$, we have

$$\begin{aligned} \mathbf{E}(\xi_u(x)) &= w \sum_{n-u \le \deg(f) < n} \mathbf{E}(X_f) e_p(x \cdot f) = w \sum_{j=n-u}^{n-1} \frac{j^c}{p^j} \sum_{\deg(f)=j} e_p(f \cdot x) \\ &= w \sum_{j=n-u}^{n-1} \frac{j^c}{p^j} \sum_{f \in G_{j+1} \setminus G_j} e_p(f \cdot x) = 0. \end{aligned} \tag{26}$$

For $n > 2^{3/c}$, we have $u \le n/2$ and $w = ((1-p^{-1})\sum_{j=n-u}^{n-1} j^c)^{-1} \le 2(u(n/2)^c)^{-1} \le 4(un^c)^{-1}$. Therefore,

$$\mathbf{Var}(\xi_u(x)) \le 2w^2 \sum_{n-u \le \deg(f) < n} \mathbf{Var}(X_f) \le 2w^2 \sum_{j=n-u}^{n-1} j^c \left(1 - \frac{j^c}{p^j}\right) \le 2w^2 u n^c \le \frac{32}{un^c}. \tag{27}$$

Moreover, for each f, $|w_u(f) e_p(f \cdot x)(X_f - \mathbf{E}(X_f))| \le 2w \le 8(un^c)^{-1}$. By Bernstein's inequality, for $\mathrm{supp}(x) \cap [0, n-u) \neq \emptyset$, we have

$$\begin{aligned} &\mathbf{P}\left(|\xi_u(x)| \ge \tfrac{1}{3}\right) \le 4\exp\left(\tfrac{-un^c}{1200}\right), \\ &\mathbf{P}\left(|\xi_u(0) - 1| \ge \tfrac{1}{3}\right) \le 4\exp\left(\tfrac{-un^c}{1200}\right). \end{aligned} \tag{28}$$

Note that $4(|S_u| + 1) \le 4(\exp(\frac{un^c}{2000}) + 1) < \exp(\frac{un^c}{1500})$ holds for all sufficiently large n depending on c. From (28), we hence can deduce that

$$\mathbf{P}\left(\left\{\max_{x \in S_u} |\xi_u(x)| \ge \frac{1}{3}\right\} \vee \left\{|\xi_u(0) - 1| \ge \frac{1}{3}\right\}\right) \le 4(|S_u|+1)\exp\left(\frac{-un^c}{1200}\right) < \exp\left(\frac{un^c}{1500}\right)\exp\left(\frac{-un^c}{1200}\right).$$

Therefore, we obtain that

$$\mathbf{P}\left(\left\{\max_{x \in S_u} |\xi_u(x)| < \frac{1}{3}\right\} \wedge \left\{|\xi_u(0) - 1| < \frac{1}{3}\right\}\right) \ge 1 - \exp\left(\frac{un^c}{1500}\right)\exp\left(\frac{-un^c}{1200}\right) = 1 - \exp\left(\frac{-un^c}{6000}\right),$$

which completes the proof. □

As promised we will now glue different ξ_u's together. The point is that we need only $O_c(1)$ of them.

Lemma 14. *Let $0 < c < 1$ and n be sufficiently large depending on c and p. Let H be the set defined in (14). There exists a (random) trigonometric polynomial*

$$\psi_n(x) = \sum_{f \in G_n} v_f e_p(f \cdot x)$$

supported on H_n with $\psi_n(0) = 1$ and

$$\mathbf{P}\left(\max_{\substack{x \in G_n,\\ x \neq 0}} |\psi_n(x)| \geq 1 - \frac{c}{12}\right) < \frac{3}{c} \exp\left(\frac{-n^c}{6000}\right). \tag{29}$$

Proof. We first take

$$u_j = \lfloor n^{1-jc/3} \rfloor \qquad \text{for } j = 1, 2, \ldots, k \tag{30}$$

where $k = \lfloor \frac{3}{c} \rfloor - 1$. Let $\xi_j(x) = \xi_{u_j}(x)$, $w_j(f) = w_{u_j}(f)$, where $\xi_{u_j}(x)$ and $w_{u_j}(f)$ are defined as in Lemma 13. Let

$$A_1 = \{x : \mathrm{supp}(x) \cap [0, n - u_1) \neq \emptyset\}. \tag{31}$$

Since $n \log p < \frac{n^{1+2c/3}}{2000}$ for sufficiently large n, we note that $|A_1| < \exp(n \log p) < \exp(\frac{u_1 n^c}{2000})$ and hence A_1 satisfies the condition of Lemma 13. In general we let

$$A_j = \{x : \mathrm{supp}(x) \subset [n - u_{j-1}, n) \text{ and } \mathrm{supp}(x) \cap [n - u_{j-1}, n - u_j) \neq \emptyset\} \tag{32}$$

for $2 \leq j \leq k$. By the definition of u_j, we note that $u_{j-1} \log p < \frac{u_j n^c}{2000}$ for large n and hence $|A_j| \leq p^{u_{j-1}} \leq \exp(\frac{u_j n^c}{2000})$. Thus, all the sets A_j satisfy the condition of Lemma 13 and we obtain

$$\mathbf{P}\left(\left\{\max_{x \in A_j} |\xi_j(x)| < \frac{1}{3}\right\} \wedge \left\{|\xi_j(0) - 1| < \frac{1}{3}\right\}\right) \geq 1 - \exp\left(\frac{-u_j n^c}{6000}\right) \tag{33}$$

for $j = 1, 2, \ldots, k$. Finally, we let

$$A_0 = (G_n \setminus \{0\}) \setminus (\cup_{j=1}^k A_j) = \{x : \mathrm{supp}(x) \subset [n - u_k, n)\}. \tag{34}$$

Since $u_k \log p < n^{2c/3} \log p < \frac{n^c}{200}$ for all sufficiently large n, $|A_0| = p^{u_k} - 1 < \exp(\frac{n^c}{200})$ holds and hence A_0 satisfies the condition of Lemma 12. Thus, for $\xi_0(x)$ defined in Lemma 12, we have

$$\mathbf{P}\left(\left\{\max_{x \in A_0} |\xi_0(x)| < \frac{1}{3}\right\} \wedge \left\{|\xi_0(0) - 1| < \frac{1}{3}\right\}\right) \geq 1 - \exp\left(\frac{-n^c}{400}\right). \tag{35}$$

We now define the trigonometric polynomial

$$\psi_n(x) := \frac{\sum_{j=0}^k \xi_j(x)}{\sum_{j=0}^k \xi_j(0)}. \tag{36}$$

Then clearly $\psi_n(0) = 1$ and ψ_n is supported on H_n because all the ξ_j are supported on H_n. Also, all the $\xi_j(0)$ are real and positive.

If all the events on the left hand sides of (33) and (35) occur, then $\sum_{j=0}^k \xi_j(0) \leq 4(k+1)/3$. If $x \in G_n \setminus \{0\}$ then there is at least one $i \in [1, k]$ such that $x \in A_i$ and consequently $|\xi_i(x)| \leq 1/3 \leq \xi_i(0) - 1/3$. For all other $j \in [1, k]$ we bound trivially $|\xi_j(x)| \leq \xi_j(0)$. Thus

$$|\psi_n(x)| = \frac{\left|\sum_{j=0}^k \xi_j(x)\right|}{\sum_{j=0}^k \xi_j(0)} \leq 1 - \frac{1/3}{\sum_{j=0}^k \xi_j(0)} < 1 - \frac{1/3}{4(k+1)/3} < 1 - \frac{c}{12}. \tag{37}$$

Consequently,

$$\mathbf{P}\left(\max_{\substack{x\in G_n,\\ x\neq 0}} |\psi_n(x)| \geq 1-\frac{c}{12}\right) \leq \mathbf{P}\Big(\Big\{\text{there exists a } j\in[1,k] \text{ s.t. (33) fails}\Big\} \vee \Big\{\text{inequality (35) fails}\Big\}\Big)$$
$$< \sum_{j=1}^{k} \exp\left(\frac{-u_j n^c}{6000}\right) + \exp\left(\frac{-n^c}{400}\right) < \frac{3}{c}\exp\left(\frac{-n^c}{6000}\right).$$

This completes the proof. □

Proof of Theorem 4. By Lemma 14, for a sufficiently large number M, we have

$$\sum_{n>M}^{\infty} \mathbf{P}\left(\max_{x\neq 0}|\psi_n(x)| \geq 1-\frac{c}{12}\right) < \sum_{n>M}^{\infty} \frac{3}{c}\exp\left(\frac{-n^c}{6000}\right) < \infty. \tag{38}$$

Therefore, by the Borel-Cantelli Lemma, the events $\{\max_{\substack{x\in G_n,\\ x\neq 0}} |\psi_n(x)| \geq 1-\frac{c}{12}\}$ occur for only finitely many n, almost surely.

Let A be any subset of $\mathbb{F}[t]$ with $\underline{d}(A) = \delta \in (0,1)$. Using Lemma 10 with $\eta = 1-\frac{c}{12}$, we obtain that

$$\liminf_{n\to\infty} \frac{|A_n+H_n|}{p^n} \geq \liminf_{n\to\infty} \left\{\frac{|A_n|}{p^n} + \left(\frac{c}{6}-\frac{c^2}{144}\right)\frac{|A_n|}{p^n}\left(1-\frac{|A_n|}{p^n}\right)\right\} \qquad \text{almost surely.} \tag{39}$$

The right-hand side of (39) is easily seen to be $\geq \delta + (\frac{c}{6}-\frac{c^2}{144})\delta(1-\delta)$, since the function $x \mapsto x+ax(1-x)$ for with $a = \frac{c}{6}-\frac{c^2}{144}$ is continuous and increasing on $(0,1)$. Thus $\liminf_{n\to\infty} \frac{|A_n+H_n|}{p^n} > \delta$ almost surely, which finishes the proof. □

4. Proof of Theorem 5

We first begin with the following Lemma, which says that if $|H_n| < n^{1+\epsilon/2}$ infinitely often, then we can find a subsequence of n such that the elements of H are well-spaced in G_n.

Lemma 15. *Suppose $H \subset G$ and $\epsilon > 0$ are such that $|H_n| < n^{1+\epsilon/2}$ infinitely often. Then there are infinitely n such that*

$$|H_n| < n^{1+\epsilon} \qquad \text{and} \qquad |H_n| - |H_m| \leq n^{\epsilon}(n-m) \qquad \text{for any } 1 \leq m \leq n. \tag{40}$$

Proof. Suppose for a contradiction that there exists $N_0 > 0$ such that for all $u > N_0$, if $|H_u| < u^{1+\epsilon}$ then there is $1 \leq v < u$ such that

$$|H_u| - |H_v| > u^{\epsilon}(u-v). \tag{41}$$

By the hypothesis, there exists $n > \max\{2N_0, 4^{1+1/\epsilon}\}$ such that

$$|H_n| < n^{1+\epsilon/2}. \tag{42}$$

Since $n > 4^{1+1/\epsilon}$, we have

$$n^{1+\epsilon/2} \leq (n/2)^{1+\epsilon}.$$

Note that for any $n/2 \leq m \leq n$, we have

$$|H_m| \leq |H_n| < n^{1+\epsilon/2} \leq (n/2)^{1+\epsilon} \leq m^{1+\epsilon}. \tag{43}$$

We apply (41) to $u = n$ and find $m_1 \in [1, n)$ such that $|H_n| - |H_{m_1}| > n^\epsilon(n - m_1)$. We put $m_0 = n$. Suppose we have found m_{i-1}. As long as $m_{i-1} \geq n/2$, thanks to (43), we can apply (41) with $u = m_{i-1}$ to find $m_i = v \in [1, m_{i-1})$. Let k be the greatest integer such that $m_{k-1} \geq n/2$, then $m_k < n/2$ and

$$|H_{m_{i-1}}| - |H_{m_i}| > m_{i-1}^\epsilon(m_{i-1} - m_i) > (n/2)^\epsilon(m_i - m_{i-1}) \tag{44}$$

for all $1 \leq i \leq k$. Summing these inequalities over $1 \leq i \leq k$, we get

$$|H_n| > (n - m_k)(n/2)^\epsilon \geq (n/2)^{1+\epsilon}. \tag{45}$$

This inequality contradicts (42). This completes the proof. □

Lemma 16. *Suppose n and H satisfy the property (40). Let $k = \lfloor \frac{1}{4\epsilon} \rfloor$. If n is sufficiently large, then there are $r_1, \ldots, r_k \in G_n$ of disjoint supports such that for any $1 \leq j \leq k$, $\mathrm{supp}(r_j) \subset [n - \lfloor \sqrt{n} \rfloor, n)$ and*

$$H_n \subset \langle r_j \rangle^\perp \cup \cap_{i=1}^{j-1} \langle r_i \rangle^\perp, \tag{46}$$

were $\langle r_i \rangle^\perp$ is the orthogonal complement in G_n of r_i. Consequently, for any $h \in H_n$, we have $h \cdot r_i = 0$ for all $i = 1, \ldots, k$ with at most one exception.

Proof. First let $d_1 := 1$ and r_1 be any vector supported on $\{n-1\}$. Since $\langle r_1 \rangle^\perp = G_{n-1}$, all elements in $H_n \setminus \langle r_1 \rangle^\perp$ are not in H_{n-1}. By inequality (40) we hence have that $\left|H_n \setminus \langle r_1 \rangle^\perp\right| \leq n^\epsilon$. Let $d_2 := \lfloor n^\epsilon \rfloor + d_1 + 2$. We shall find r_2 with $\mathrm{supp}(r_2) \subset [n - d_2, n - d_1)$ such that $H_n \setminus \langle r_1 \rangle^\perp \subset \langle r_2 \rangle^\perp$. The subspace $\langle H_n \setminus \langle r_1 \rangle^\perp \rangle^\perp$ has dimension at least $n - \lfloor n^\epsilon \rfloor - 1$ and the subspace spanned by $\{t \in G_n : \mathrm{supp}(t) \subset [n - d_2, n - d_1)\}$ has dimension $d_1 - d_2 = \lfloor n^\epsilon \rfloor + 2$. The sum of these dimensions is greater than n, which implies that the two subspaces has nonzero intersection. Thus we can find a vector r_2 supported on $[n - d_2, n - d_1)$ satisfying $h \cdot r_2 = 0$ for all $h \in H_n \setminus \langle r_1 \rangle^\perp$.

In general, suppose we have found $\{r_i\}_{i=1}^{j-1}$ and $\{d_i\}_{i=1}^{j-1}$ such that $\mathrm{supp}(r_i) \in [n - d_i, n - d_{i-1})$. We next want to find r_j satisfying

$$H_n \setminus \cap_{i=1}^{j-1} \langle r_i \rangle^\perp \subset \langle r_j \rangle^\perp. \tag{47}$$

Since $H_n \setminus \cap_{i=1}^{j-1} \langle r_i \rangle^\perp$ is supported on $[0, n - d_{j-1})$, by property (40), we have $|H_n \setminus \cap_{i=1}^{j-1} \langle r_i \rangle^\perp| \leq n^\epsilon d_{j-1}$ and hence $\langle H_n \setminus \cap_{i=1}^{j-1} \langle r_i \rangle^\perp \rangle^\perp$ has dimension at least $n - \lfloor n^\epsilon d_{j-1} \rfloor - 1$. Further by letting

$$d_j := \lfloor n^\epsilon d_{j-1} \rfloor + d_{j-1} + 2, \tag{48}$$

the dimension of the subspace spanned by $\{t \in G_n : \mathrm{supp}(t) \subset [n - d_j, n - d_{j-1})\}$ is $d_j - d_{j-1} = \lfloor n^\epsilon d_{j-1} \rfloor + 2$. Thus the sum of the dimensions of these two subspaces is greater than n and their intersection must be nonzero, which yields a r_j such that $\mathrm{supp}(r_j) \in (n - d_j, n - d_{j-1}]$ and $h \cdot r_j = 0$ for all $h \in H_n \setminus \cap_{i=1}^{j-1} \langle r_i \rangle^\perp$.

We can continue this process as long as $d_j < n$. From (48) we obtain that $d_j \leq (n^\epsilon + 3) d_{j-1}$ for all j. For $k = \lfloor \frac{1}{4\epsilon} \rfloor$, we have

$$d_k \leq (n^\epsilon + 3)^k < n^{2\epsilon k} < \lfloor \sqrt{n} \rfloor < n,$$

which means that we can construct k vectors $\{r_j\}_{j=1}^k$ of disjoint supports and $\mathrm{supp}(r_j) \subset (n - \lfloor \sqrt{n} \rfloor, n)$ for all $j = 1, \cdots, k$.

Now it remains to show that for every $h \in H_n$, $h \cdot r_j = 0$ holds for all $1 \leq j \leq k$ with at most one exception. On rewriting (47), we obtain the formula (46) for all $1 \leq j \leq k$. Take $h \in H_n$ and let ℓ be the first index such that $h \notin \langle r_\ell \rangle^\perp$. If $\ell = k$, then r_k could be the exception. If $\ell < k$, by taking $\ell \leq j \leq k$ in (46), we know h has to be in $\langle r_i \rangle^\perp$ for all $\ell + 1 \leq i \leq k$, in which case r_ℓ is the exception. This completes the proof. □

Proposition 17. *Let $0 < \delta < 1$ and $\epsilon > 0$. Suppose $H \subset G$ is such that for any $\epsilon > 0$, $|H_n| < n^{1+\epsilon/2}$ infinitely often. Then for each sufficiently large n satisfying (40), there exists a subset B_n satisfying the following four properties:*

(i) $\delta \leq \frac{|B_n|}{p^n}$;
(ii) $\frac{|B_n+H_n|}{p^n} \leq \delta + O(\epsilon^{1/2})$;
(iii) $\frac{|B_n \cap G_m|}{p^m} \geq \frac{|B_n|}{p^n}$ *for all* $0 \leq m \leq n$;
(iv) $G_{n-\lfloor\sqrt{n}\rfloor} \subset B_n$.

Proof. Let $k = \lfloor \frac{1}{4\epsilon} \rfloor$. For any sufficiently large n satisfying (40), let $\{r_j\}_{j=1}^k$ be vectors of disjoint supports and supported on $(n - \lfloor\sqrt{n}\rfloor, n)$ given by Lemma 16.

For $f \in G_n$, we define $X_j(f) = \mathrm{Re}(e_p(f \cdot r_j))$. Since r_j is supported on $(n - \lfloor\sqrt{n}\rfloor, n)$, X_j is constant on translates of $G_{n-\lfloor\sqrt{n}\rfloor}$. Since the r_j's have disjoint supports, we can regard the X_j's as independent random variables from G_n to $\mathbb{R}$. It is easy to see that

$$\mathbf{E}(X_j) = 0, \qquad \mathbf{Var}(X_j) = \begin{cases} 1/2, & \text{if } p \neq 2 \\ 1, & \text{if } p = 2 \end{cases} \qquad \text{and} \qquad \mathbf{E}(|X_j - \mathbf{E}(X_j)|^3) \leq 1. \tag{49}$$

Now we define

$$X = \sum_{j=1}^{k} X_j \tag{50}$$

and

$$F(x) = \begin{cases} \mathbf{P}(\sqrt{2/k}X \leq x) & \text{if } p \neq 2 \\ \mathbf{P}(\sqrt{1/k}X \leq x) & \text{if } p = 2. \end{cases}$$

By the Berry-Esseen inequality (Lemma 7), we have

$$\sup_{x \in \mathbb{R}} |F(x) - \Phi(x)| \leq \frac{2\sqrt{2}}{\sqrt{k}} \tag{51}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. For each $m \in \mathbb{Z}$, define the niveau set

$$S_m = \{f : f \in G_n, X(f) \geq m\}. \tag{52}$$

Then $G_{n-\lfloor\sqrt{n}\rfloor} = S_k \subset S_{k-1} \subset \cdots$. Since X is constant on translates of $G_{n-\lfloor\sqrt{n}\rfloor}$, if $x \in S_m$, then $x + G_{n-\lfloor\sqrt{n}\rfloor} \subset S_m$.

For any $h \in H_n$ and $f \in G_n$, we have

$$|X(f+h) - X(f)| = \left| \sum_{j=1}^{k} \mathrm{Re}(e_p(f \cdot r_j)(e_p(h \cdot r_j) - 1)) \right| \leq 2 \tag{53}$$

since $h \cdot r_j = 0$ with at most one exception. From the definition of S_m, this implies that

$$S_m + H_n \subset S_{m-2} \tag{54}$$

for any m.

Let M be the largest integer such that $|S_M| \geq \delta p^n$, then $M < k$ if n is sufficiently large. We let $B_n = S_M$. By the definition of M, we have $|S_{M+1}| < \delta p^n$ and $G_{n-\lfloor\sqrt{n}\rfloor} \subset B_n$.

From (54) we have $B_n + H_n \subset S_{M-2}$ and

$$\begin{aligned}\frac{|B_n + H_n|}{|G_n|} &\le \frac{|S_{M+1}|}{|G_n|} + \frac{|S_{M-2} \setminus S_{M+1}|}{|G_n|} \\ &\le \delta + \frac{|\{f \in G_n : M-2 \le X(f) < M+1\}|}{|G_n|} \\ &= \begin{cases} \delta + F(\sqrt{2/k}(M+1)) - F(\sqrt{2/k}(M-2)) & \text{if } p \neq 2, \\ \delta + F(\sqrt{1/k}(M+1)) - F(\sqrt{1/k}(M-2)) & \text{if } p = 2. \end{cases}\end{aligned} \tag{55}$$

The triangle inequality and (51) imply that for all $a > b$

$$|F(a) - F(b)| \le |\Phi(a) - \Phi(b)| + 4\sqrt{2/k}. \tag{56}$$

Further, we note that

$$|\Phi(a) - \Phi(b)| = \frac{1}{\sqrt{2\pi}} \left| \int_a^b e^{-u^2/2} du \right| \le |a - b|. \tag{57}$$

Combining this inequality with (56) and (55), we have

$$\frac{|B_n + H_n|}{p^n} \le \delta + 7\sqrt{2/k} = \delta + O(\sqrt{\epsilon}). \tag{58}$$

Recall that by Lemma 11, there exists a vector $x_n \in B_n$ such that $\frac{|(B_n - x_n) \cap G_m|}{p^m} \ge \frac{|B_n|}{p^n}$ for all $0 \le m \le n$. Since $G_{n-\lfloor\sqrt{n}\rfloor} \subset B_n - x_n$, Proposition 17 follows by taking the shifted set as our new B_n. □

Proof of Theorem 5. Fix $0 < \delta < 1$, and suppose that for any $\epsilon > 0$, $|H_n| < n^{1+\epsilon}$ holds for infinitely many n. By Lemma 15, for each $k > 1$, there are infinitely many n such that $|H_n| < n^{1+1/k}$ and (40) holds with $\epsilon = 1/k$. Let n_k be such an n, and since there are infinitely many choices for n_k, we may require that $n_k - \lfloor\sqrt{n_k}\rfloor > 2n_{k-1}$ for any $k > 0$.

Let $B_{n_k} \subset G_{n_k}$ be the set provided by Proposition 17 with $\epsilon = 1/k$. Our goal is to glue the sets B_{n_k} together. Set

$$A := \bigcup_{k=1}^{\infty} (B_{n_k} \setminus G_{n_{k-1}}) \tag{59}$$

where we define $G_{n_0} = \emptyset$. (A simple union $\cup_{k=1}^{\infty} B_{n_k}$ won't work; this is where our construction differs from Ruzsa's.) Note that by Proposition 17 (iv), $B_{n_k} \supset G_{2n_{k-1}} \supset G_{n_{k-1}}$, so the union in (59) is a disjoint union.

For any $m > 0$, we have

$$\begin{aligned} A_m &= \bigcup_{n_l \ge m} \left(G_m \cap (B_{n_{l+1}} \setminus G_{n_l})\right) \cup \bigcup_{n_l < m} \left(G_m \cap (B_{n_{l+1}} \setminus G_{n_l})\right) \\ &= \bigcup_{n_l < m} \left(G_m \cap (B_{n_{l+1}} \setminus G_{n_l})\right). \end{aligned} \tag{60}$$

Claim 1: $\liminf_{n\to\infty} \frac{|A_n + H_n|}{|G_n|} \le \delta$.

Indeed, from (60) we have $A_{n_k} = \cup_{l=1}^{k} \left(G_{n_k} \cap (B_{n_l} \setminus G_{n_{l-1}})\right) \subset \cup_{l=1}^{k} (G_{n_k} \cap B_{n_l})$ and

$$\begin{aligned} \frac{|A_{n_k} + H_{n_k}|}{|G_{n_k}|} &\leq \frac{|B_{n_k} + H_{n_k}|}{|G_{n_k}|} + \sum_{l=1}^{k-1} \frac{|B_{n_l} + H_{n_k}|}{|G_{n_k}|} \\ &\leq \delta + O(\epsilon_k^{-1/2}) + \frac{\sum_{l=1}^{k-1} n_k^{1+1/k} p^{n_l}}{p^{n_k}} \\ &\leq \delta + O(\epsilon_k^{-1/2}) + O(n_k^{1+1/k} p^{-n_k/2}) \end{aligned}$$

where on the second line we use Proposition 17 (ii) and the trivial bound $|B_{n_l} + H_{n_k}| \leq |H_{n_k}||B_{n_l}| \leq |H_{n_k}| p^{n_l}$. Letting $k \to \infty$, the claim follows.

Claim 2: $\liminf_{n\to\infty} \frac{|A_n|}{p^n} \geq \delta$. Indeed, we will show that for any m with $n_k < m \leq n_{k+1}$, we have

$$\frac{|A_m|}{p^m} \geq \delta - \frac{1}{p^{n_{k-1}}}. \tag{61}$$

We distinguish two cases:

Case 1: When $2n_k < m \leq n_{k+1}$, from (60) we have

$$\begin{aligned} \frac{|A_m|}{p^m} &\geq \frac{|(B_{n_{k+1}} \setminus G_{n_k}) \cap G_m|}{p^m} = \frac{|(B_{n_{k+1}} \cap G_m) \setminus G_{n_k}|}{p^m} \\ &\geq \frac{|(B_{n_{k+1}} \cap G_m)| - |G_{n_k}|}{p^m} \\ &\geq \delta - \frac{1}{p^{m-n_k}} \geq \delta - \frac{1}{p^{n_k}}, \end{aligned} \tag{62}$$

by Proposition 17 (i), (iii), and the fact that $m \geq 2n_k$.

Case 2: When $n_k < m \leq 2n_k$, then again from (60) we have

$$A_m \supset \left((B_{n_{k+1}} \cap G_m) \setminus G_{n_k}\right) \cup \left((B_{n_k} \cap G_m) \setminus G_{n_{k-1}}\right) = (G_m \setminus G_{n_k}) \cup (B_{n_k} \setminus G_{n_{k-1}}),$$

where we have used the fact that $B_{n_k} \subset G_{n_k} \subset G_m \subset G_{2n_k} \subset B_{n_{k+1}}$. Hence,

$$\begin{aligned} \frac{|A_m|}{p^m} &\geq 1 - \frac{1}{p^{m-n_k}} + \frac{\delta}{p^{m-n_k}} - \frac{1}{p^{m-n_{k-1}}} \\ &\geq \delta - \frac{1}{p^{n_{k-1}}}, \end{aligned} \tag{63}$$

since $m > n_k > 2n_{k-1}$ and $1 - \frac{1}{a} + \frac{\delta}{a} \geq \delta$ for $a := p^{m-n_k} > 1$. Thus in any case (61) is true, and $\liminf_{n\to\infty} \frac{|A_n|}{|G_n|} \geq \delta$.

Putting everything together, we have

$$\delta \leq \liminf_{n\to\infty} \frac{|A_n|}{|G_n|} \leq \liminf_{n\to\infty} \frac{|A_n + H_n|}{|G_n|} \leq \delta, \tag{64}$$

which implies $\delta = \underline{d}(A) = \liminf_{n\to\infty} \frac{|A_n + H_n|}{|G_n|}$, as desired. □

5. Construction of an explicit essential component

Recall (Remark 2) that a multiset $S \subset G_n$ is ϵ-biased if

$$\forall c \in \mathbb{F}, \forall x \in G_n, x \neq 0, \qquad \left| \frac{1}{|S|} \#\{s \in S : x \cdot s = c\} - \frac{1}{p} \right| \leq \epsilon.$$

Lemma 10 implies that if $S \subset G_n$ is ϵ-biased, then its underlying set K is an essential component of G_n. In theoretical computer science, it is desirable to construct such a multiset S that has a small size relative to both n and ϵ. The current record (at least when $p = 2$) is due to Ta-Shma [17], who constructed a multiset S with $|S| = O(\frac{n}{\epsilon^{2+o(1)}})$. For our purpose, we only need to work with any small, fixed ϵ (say $\epsilon = \frac{1}{2p}$), so the dependence on ϵ is unimportant.

It turns out that we cannot simply use constructions of ϵ-biased sample spaces as a "blackbox". Naturally, in order to construct an essential component H in $\mathbb{F}[t]$, one would like to take H to be the union of K_m, where K_m is an essential component of G_m. This, however, does not guarantee that H has small counting function, since $H_n = G_n \cap (\bigcup_{m=1}^{\infty} K_m) = \bigcup_{m=1}^{\infty}(K_m \cap G_n)$, and K_m may have nonempty intersection with G_n for $m > n$. Thus one needs information on the supports of elements of K_m. Alon-Goldreich-Håstad-Peralta's construction [1, Section 3] (see also [6, Theorem 2] for an exposition) is very simple and suits well our purpose. This construction gives $|S| = O_p(n^2)$, which is why our essential component has counting function $O_p(n^3)$. We will now describe their construction and also sketch the proof for the sake of completeness (Alon-Goldreich-Håstad-Peralta only worked with $p = 2$, but the construction works for any p).

Let $\ell = \lfloor \log_p n + C_p \rfloor$ for some constant C_p. Let P_ℓ be the set of all irreducible polynomials polynomials in $\mathbb{F}[t]$ with degree ℓ and leading coefficient -1. For each $s = (s_0, \ldots, s_{\ell-1}) \in G_\ell$ and $f = (f_0, f_1, \ldots, f_{\ell-1}, -1) \in P_\ell$, we define an element $r = r(s, f) = (r_0, \ldots, r_{n-1}) \in G_n$ as follows

$$r_i = \begin{cases} s_i & \text{for } 0 \leq i \leq \ell - 1, \\ \sum_{j=0}^{\ell-1} f_j r_{i-\ell+j} & \text{for } \ell \leq i \leq n-1. \end{cases}$$

Claim: The multiset $S = \{r(s, f) : s \in G_\ell, f \in P_\ell\}$ is $O_p(\frac{n}{p^\ell})$-biased.

By adjusting C_p, we can make the quantity $O_p(\frac{n}{p^\ell})$ less than $\frac{1}{2p}$. Clearly $|S| \leq p^{2\ell} = O_p(n^2)$.

It remains to prove the claim. Let us fix $x \in G_n \setminus \{0\}$ and $c \in \mathbb{F}$. We want to estimate $\mathbf{P}_{s \in G_\ell, f \in P_\ell}(x \cdot r(s, f) = c)$. Without loss of generality we may assume $c \neq 0$. For each fixed $f \in P_\ell$, the map $s \mapsto r(s, f)$ from G_ℓ to G_n is linear, and we denote its matrix by M_f. We have

$$\mathbf{P}_{s \in G_\ell, f \in P_\ell}(x \cdot r(s, f) = c) = \mathbf{P}_{s \in G_\ell, f \in P_\ell}(x \cdot M_f s = c) = \mathbf{P}_{s \in G_\ell, f \in P_\ell}(M_f^T x \cdot s = c)$$

where M_f^T is the transpose of M_f. For each f, we have $\mathbf{P}_{s \in G_\ell}(M_f^T x \cdot s = c)$ is exactly $\frac{1}{p}$ if $M_f^T x \neq 0$, and 0 if $M_f^T x = 0$. Thus the probability above is equal to $\frac{1}{p}(1 - \mathbf{P}_{f \in P_\ell}(M_f^T x = 0))$.

On the other hand, by the construction of M_f, we can see that $M_f^T x$ is actually the reduction of x modulo f (it suffices to verify this for $x = t^i, 1 \leq i \leq n-1$). Hence if $M_f^T x = 0$, then f divides x. But x cannot have more than $\frac{n}{\ell}$ irreducible factors of degree ℓ. Therefore, $\mathbf{P}_{f \in P_\ell}(M_f^T x = 0) \leq \frac{n/\ell}{|P_\ell|} = O_p(\frac{n}{p^\ell})$.

We are now ready to prove Theorem 6.

Proof of Theorem 6. For each m, let $S_m \subset G_m$ be the $\frac{1}{2p}$-biased set given by the construction above and K_m be its underlying set. In particular $|K_m| \leq |S_m| = O_p(m^2)$. We now define

$$H = \bigcup_{m=0}^{\infty} (K_m - t^{m-1}).$$

Let $A \subset G$ be a subset with $\underline{d}(A) = \delta \in (0,1)$. Then for any n sufficiently large, we have

$$\frac{|A_n + H_n|}{p^n} \geq \frac{|A_n + K_n - t^{n-1}|}{p^n} = \frac{|A_n + K_n|}{p^n} \geq \frac{|A_n|}{p^n} + c\frac{|A_n|}{p^n}\left(1 - \frac{|A_n|}{p^n}\right)$$

for some constant $c \in (0,1)$. Taking lim inf of both sides, we have

$$\liminf_{n\to\infty} \frac{|A_n + H_n|}{p^n} \geq \liminf_{n\to\infty} \left(\frac{|A_n|}{p^n} + c\frac{|A_n|}{p^n}\left(1 - \frac{|A_n|}{p^n}\right)\right) \geq \delta + c\delta(1-\delta)$$

since the function $x \mapsto x + cx(1-x)$ is increasing on (0,1). This shows that H is an essential component in G.

It remains to estimate $|H_n|$. We have

$$H_n = H \cap G_n = \bigcup_{m=1}^{\infty} ((K_m - t^{m-1}) \cap G_n).$$

Claim: If $(K_m - t^{m-1}) \cap G_n \neq \emptyset$ and n is sufficiently large, then $m \leq 2n$.

Indeed, suppose for a contradiction that $m > 2n$. Let $x = (x_0, x_1, \ldots, x_{m-1}) \in K_m \cap (t^{m-1} + G_n)$. Then $x_{m-1} = 1$, while $x_i = 0$ for any $n \leq i < m-1$. By the construction of S_m, we have $x = r(f,s)$ for some $s \in G_\ell$ and $f \in P_\ell$, where $\ell = \lfloor \log_p m + C_p \rfloor$. If n is sufficiently large, then $m - 1 - \ell \geq n$. This yields the desired contradiction since $1 = x_{m-1} = \sum_{j=0}^{\ell-1} f_j x_{m-1-\ell+j} = 0$.

Hence, we have

$$|H_n| = \left|\bigcup_{m=1}^{\infty} ((K_m + t^{m-1}) \cap G_n)\right| \leq \sum_{m=0}^{2n} |K_m| = O_p(n^3)$$

as desired. □

By using a similar idea, and by using an isoperimetric inequality in $\mathbb{F}^n$ ([5, Theorem 1.2]) one can prove that for any $\eta > 0$, the set

$$H = \cup_{n=1}^{\infty} \left\{x + \mathbf{1}_n : x \in G_n, |\mathrm{supp}(x)| \leq \eta\sqrt{n}\right\}$$

is an essential component in G, where $\mathbf{1}_n := 1 + t + \cdots + t^{n-1}$. This essential component has the advantage of being simpler, but its counting function is $|H_n| = \exp\left(O_p\left(\eta\sqrt{n}\log n\right)\right)$. This set is the analog of Wirsing's example (1).

Erdős [16, p. 147] asked whether the set $\{2^n 3^m : m, n \in \mathbb{N}\}$ is an essential component in $\mathbb{N}$. This is in keeping with the principle that multiplicative and additive structures don't mix well, as exemplified by sum-product estimates. Note that the counting function of this set is $O(\log^2 x)$. Erdős' question remains open. The following question is perhaps more tractable.

Problem. *Can one prove or disprove a similar statement in $\mathbb{F}_p[t]$? For example, is the set $\{t^n(t+1)^m : m, n \in \mathbb{N}\}$ an essential component in $\mathbb{F}_2[t]$?*

References

[1] N. Alon, O. Goldreich, J. Håstad, R. Peralta, *Simple constructions of almost k-wise independent random variables*, Random Structures Algorithms 3 (1992), no. 3, 289–304.

[2] S. Boucheron, G. Lugosi, and P. Massart, **Concentration inequalities: A nonasymptotic theory of independence**, Oxford University Press, (2016).

[3] J. R. Burke, *A notion of density and essential components in $GF[p, x]$*, Acta Arith. 44 (1984), no. 4, 299–306.

[4] P. Erdős, *On the arithmetical density of the sum of two sequences one of which forms a basis for the integers*, Acta. Arith., 1(2) (1935), 197–200.

[5] Z. Ge, T. H. Lê, *On theorems of Wirsing and Sanders*, Acta Arithmetica, 189 (2019), 381–390.

[6] O. Goldreich, Randomized Methods in Computation, Lecture 8: Small Bias Sample Spaces, http://www.wisdom.weizmann.ac.il/∼oded/PS/RND/l08.ps

[7] H. Halberstam, K. F. Roth, **Sequences**. Springer-Verlag, New York, 1983.

[8] A. Khinchin, *Über ein metrisches Problem der additiven Zahlentheorie*, Mat. Sb., 40:2 (1933), 180–189.

[9] B. Green, *Some constructions in the inverse spectral theory of cyclic groups*, Combin. Probab. Comput. 12 (2003), no. 2, 127–138.

[10] T. H. Lê, **Topics in arithmetic combinatorics in function fields**, PhD Thesis, UCLA (2010).

[11] U. V. Linnik, *On Erdős's theorem on the addition of numerical sequences*, Rec. Math. [Mat. Sbornik] N.S., 10(52):1–2 (1942), 67–78.

[12] H. Plünnecke, **Eigenschaften und Abschdtzungen von Wirkungsfunktionen**, Ges. Mathem. und Datenverarbeitung 22 (Bonn, 1969).

[13] I. Z. Ruzsa, *Sumsets and structure*, **Combinatorial number theory and additive group theory**, Advanced Courses in Mathematics, CRM Barcelona, Birkhäuser Verlag, Basel, 2009.

[14] I. Z. Ruzsa, *Essential components*, Proc. London Math. Soc. (3) 54 (1987), no. 1, 38–56.

[15] I. Z. Ruzsa, *Arithmetic progressions in sumsets*, Acta Arith. 60 (1991), no. 2, 191–202.

[16] I. Z. Ruzsa, *Erdős and the integers*, J. Number Theory 79 (1999), no. 1, 115–163.

[17] A. Ta-Shma, *Explicit, almost optimal, epsilon-balanced codes*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (2017), 238–251.

[18] E. Wirsing, *Thin essential components*, **Topics in number theory** (Proc. Colloq., Debrecen, 1974), pp. 429–442. Colloq. Math. Soc. János Bolyai, Vol. 13, North-Holland, Amsterdam, 1976.

[19] J. Wolf, *The structure of popular difference sets*, Israel J. Math. 179 (2010), 253–278.

[20] A. Gut, *Probability: A Graduate Course (Springer Texts in Statistics)*, Springer-Verlag New York, (2013)

[21] Y. V. Prokhorov, V. Statulevicius (Eds.), *Limit Theorems of Probability Theory*, Springer-Verlag Berlin Heidelberg, (2000)

Z. Ge, Institute of Mathematical Sciences, ShanghaiTech University, Shanghai, China

Email address: gezc@shanghaitech.edu.cn

T. H. Lê, Department of Mathematics, The University of Mississippi, University, MS 38677

Email address: leth@olemiss.edu