ON CONSTANT-MULTIPLE-FREE SETS CONTAINED IN A RANDOM SET OF INTEGERS

SANG JUNE LEE

Abstract. For a rational number \( r > 1 \), a set \( A \) of positive integers is called an \( r \)-multiple-free set if \( A \) does not contain any solution of the equation \( rx = y \). The extremal problem on estimating the maximum possible size of \( r \)-multiple-free sets contained in \( [n] := \{1, 2, \ldots, n\} \) has been studied for its own interest in combinatorial number theory and for application to coding theory. Let \( a, b \) be positive integers such that \( a < b \) and the greatest common divisor of \( a \) and \( b \) is 1. Wakeham and Wood showed that the maximum size of \((b/a)\)-multiple-free sets contained in \([n]\) is \( \frac{b}{b+a} n + O(\log n) \).

In this note we generalize this result as follows. For a real number \( p \in (0, 1) \), let \([n]_p\) be a set of integers obtained by choosing each element \( i \in [n] \) randomly and independently with probability \( p \). We show that the maximum possible size of \((b/a)\)-multiple-free sets contained in \([n]_p\) is \( \frac{b}{b+a} pm + O(\sqrt{pn} \log n \log \log n) \) with probability that goes to 1 as \( n \to \infty \).

1. Introduction

In recent years a trend in extremal combinatorics concerned with investigating how classical extremal results in dense environments transfer to sparse settings, and it has seen to be a fruitful subject of research. Especially, in combinatorial number theory, the following extremal problem in a dense environment has been well-studied and successively extended to sparse settings: Fix an equation and estimate the maximum size of subsets of \([n] := \{1, 2, \ldots, n\} \) containing no non-trivial solutions of the given equation.

An example of this line of research is a version of Roth’s theorem \([10]\) on arithmetic progressions of length 3 (with respect to the equation \( x_1 + x_3 = 2x_2 \)) for random subsets of integers in Kohayakawa–Luczak–Rödl \([8]\). Also, Szemerédi’s theorem \([12]\) was transferred to random subsets of integers in Conlon–Gowers \([2]\) and Schacht \([11]\). The result of Erdős–Turán \([4]\), Chowla \([1]\), and Erdős \([3]\) in 1940s on the maximum size of Sidon sets in \([n]\) was extended in \([6, 7]\) to sparse random subsets of \([n]\), where a Sidon...
**set** is a set of positive integers not containing any non-trivial solution of \( x_1 + x_2 = y_1 + y_2 \).

In this paper we transfer the following extremal results to sparse random subsets. For a rational number \( r > 1 \), a set \( A \) of positive integers is called an \( r \)-multiple-free set if \( A \) does not contain any solution of \( rx = y \). An interesting problem on \( r \)-multiple-free sets is of estimating the maximum possible size \( f_r(n) \) of \( r \)-multiple-free sets contained in \([n] := \{1, 2, \ldots, n\}\). This extremal problem has been studied in [14, 9, 13] for its own interest in combinatorial number theory, and also was applied to coding theory in [5].

Wang [14] showed that \( f_2(n) = \frac{2}{3}n + O(\log n) \). Leung and Wei [9] proved that for every integer \( r > 1 \), \( f_r(n) = \frac{r}{r+1}n + O(\log n) \). Wakeham and Wood [13] extended it to rational numbers as follows. For positive integers \( a \) and \( b \), let \( \gcd(b, a) \) be the greatest common divisor of \( a \) and \( b \).

**Theorem 1** (Wakeham and Wood [13]). Let \( a, b \) be positive integers with \( a < b \) and \( \gcd(b, a) = 1 \). Then

\[
 f_{b/a}(n) = \frac{b}{b + 1}n + O(\log n).
\]

We shall investigate the maximum size of constant-multiple-free sets contained in a random subset of \([n]\). Let \([n]_p \) be a random subset of \([n]\) obtained by choosing each element in \([n]\) independently with probability \( p \). Let \( f_r([n]_p) \) denote the maximum size of \( r \)-multiple-free sets contained in \([n]_p \). We are interested in the behavior of \( f_r([n]_p) \) for every rational number \( r > 1 \).

Theorem 1 gives the answer of the above question for the case \( p = 1 \). On the other hand, if \( p = o(1) \), then the usual deletion methods give that with high probability (that is, with probability that goes to 1 as \( n \to \infty \)) the maximum size of \((b/a)\)-multiple-free sets contained in \([n]_p \) is \( np(1 - o(1)) \). Hence, from now on, we consider \( p \) as a real number with \( 0 < p < 1 \).

Using Chernoff bounds (for example, see Lemma 11), Theorem 1 easily implies the following:

**Fact 2.** Let \( p \in (0, 1) \) and let \( a, b \) be positive integers such that \( a < b \) and \( \gcd(a, b) = 1 \). Let \( \omega \) be a function of \( n \) that goes to \( \infty \) arbitrarily slowly as \( n \to \infty \). With high probability, there is a \((b/a)\)-multiple-free set in \([n]_p \) of size

\[
 \frac{b}{b + 1}pn + \omega \sqrt{pn}.
\]

Fact 2 gives a lower bound on \( f_{b/a}([n]_p) \) that is off from the right value of \( f_{b/a}([n]_p) \). The main result of this paper is the following:
Theorem 3. Let $p \in (0,1)$ and let $a,b$ be positive integers such that $a < b$ and $\gcd(a,b) = 1$. Then, with high probability,

$$f_{b/a}([n]_p) = \frac{b}{b+p}pn + O\left(\sqrt{pn \log n \log \log n}\right).$$

The ratio $\frac{f_{b/a}([n]_p)}{np}$ goes from 1 to $\frac{b}{b+1}$ as $p$ varies from 0 to 1 (See Figure 1). The proof of Theorem 3 is given in Sections 2 and 3 by using a graph theoretic method.

2. Proof of Theorem 3

In order to show Theorem 3, we use a graph theoretic approach that was used in Wakeham and Wood [13]. Let $r = \frac{b}{a} > 1$ be a rational number. Let $D = (V,E)$ be the directed graph with the vertex set $V = [n]$ in which the set $E$ of arcs (or directed edges) is $\{(x,y): rx = y\}$. Let $D([n]_p)$ be the subgraph of $D$ induced on $[n]_p$. Observe that $f_r([n]_p)$ is the same as the independence number $\alpha(D([n]_p))$ of $D([n]_p)$.

We consider structures of $D([n]_p)$. The indegree and outdegree of each vertex in $D$ are at most 1. Also, there is no directed cycle in $D$ because $(x,y) \in E$ implies $x < y$. Therefore, each component of $D$ or $D([n]_p)$ is a directed path.

In order to obtain an independent set of $D([n]_p)$ of maximum size, we consider such an independent set componentwise. Let $C$ be a component of $D([n]_p)$. As we mentioned above, $C$ is a directed path. Let $V(C) = \{u_0, u_1, u_2, \cdots, u_i, \cdots, u_l\}$ be the vertex set of $C$ such that $u_j < u_{j+1}$ and $(u_j, u_{j+1}) \in E$ for $0 \leq j \leq l-1$. Observe that $V^*(C) := \{u_0, u_2, u_4, \cdots\} \subset$
$V(C)$ forms an independent set of $C$ of maximum size. Therefore, the set

$$T^* := \bigcup_C V^*(C),$$

where $C$ is each component of $D[[n]_p]$, forms an independent set of $D[[n]_p]$ of maximum size. Hence, we have the following.

**Lemma 4.** $f_r([n]_p) = |T^*|.$

Thus, in order to show Theorem 3, it suffices to show the following.

**Lemma 5.** Let $p \in (0,1)$ and let $a, b$ be natural numbers such that $a < b$ and $\gcd(a, b) = 1$. Then, with high probability,

$$|T^*| = \frac{b}{b + p}pn + O \left( \sqrt{pn \log n \log \log n} \right).$$

The proof of Lemma 5 is given in Section 3.

### 3. Proof of Lemma 5

From now on, we show Lemma 5. For positive integers $b$ and $k$, let $k$ be an $i$-th subpower of $b$ if $k = b^i l$ for some $l \not\equiv 0 \pmod{b}$. Let $T_i$ be the set of $i$-th subpowers of $b$ in $[n]$. Let $T_i^* \subset T_i$ denote the set of $i$-th subpowers $v$ of $b$ in $[n]_p$ such that $v$ is at an even distance from the smallest vertex of the component of $D[[n]_p]$ containing $v$. Observe that $T^* = \bigcup_i T_i^*$, and hence,

$$|T^*| = \sum_i |T_i^*|. \quad (1)$$

In Section 3.1 we estimate the expected value $\mathbb{E}(|T_i^*|)$. Section 3.2 deals with a concentration result of $|T^*|$ with high probability.

#### 3.1. Expectation

We first estimate $\mathbb{E}(|T_i^*|)$ and their sum $\mathbb{E}(|T^*|)$. Recall that $T_i$ denotes the set of $i$-th subpowers of $b$ in $[n]$. Note that since $1 \leq b^i \leq n$, the range of $i$ is $0 \leq i \leq \log_b n$. It is clear that

$$T_i = \left\{ b^i x \mid 1 \leq x \leq \frac{n}{b^i}, \quad x \not\equiv 0 \pmod{b} \right\}. $$

Hence we have the following:

**Fact 6.**

$$|T_i^*| = \frac{b - 1}{b^{1+i}} n \pm 1. \quad (2)$$

We consider two cases separately, based on the parity of $i$.

**Lemma 7.** For $0 \leq j \leq (\log_b n)/2$, we have

$$\mathbb{E}(|T^*_i|) = \frac{b - 1}{b(1+p)}pn \left( \frac{1}{b^{2j}} + \left( \frac{p}{b} \right)^{2j} p \right) \pm 1.$$
Proof. First we consider $\Pr \left[ v \in T_{2j} \right]$. Let $\{v_0, v_1, v_2, \cdots\}$, where $v_i < v_{i+1}$, be the vertex set of the component of $D$ containing $v$. Observe that $v_i \in T_i$, and hence, $v = v_{2j}$. The event that $v \in T_{2j}$ is in $T^*_{2j}$ happens only when one of the following holds:

- There is some $r$ with $0 \leq r \leq j - 1$ such that $v_{2j-1-2r} \notin [n]_p$ and $v_i \in [n]_p$ for all $2j - 2r \leq i \leq 2j$.
- The vertices $v_0, v_1, \cdots, v_{2j}$ are in $[n]_p$.

Hence, we have

$$\Pr \left[ v \in T_{2j} \right] = p \left( 1 - p \right) + p^2 \left( 1 - p \right) + \cdots + p^{2j-2} \left( 1 - p \right) + p^{2j}.$$  

Thus we infer

$$\mathbb{E} \left( |T^*_{2j}| \right) = |T_{2j}| \cdot \Pr \left[ v \in T_{2j} \right]$$

$$\mathbb{E} \left( |T^*_{2j}| \right) = \frac{b-1}{b(1+p)} p \left( 1 - p \right) \left( 1 - p \right)^{\frac{b}{2j}} + p^{2j}$$

$$\mathbb{E} \left( |T^*_{2j}| \right) = \frac{b-1}{b(1+p)} p \left( 1 - p \right) \left( 1 - p \right)^{\frac{b}{2j}} + p^{2j}$$

which completes the proof of Lemma 7.

Lemma 8. For $1 \leq j \leq (\log_b n)/2$, we have

$$\mathbb{E} \left( |T^*_{2j-1}| \right) = \frac{b-1}{b(1+p)} p \left( 1 - p \right) \left( 1 - p \right)^{\frac{b}{2j-1}} + p^{2j-1} \pm 1.$$  

Proof. Using an argument similar to the proof of (3), one may obtain that

$$\Pr \left[ v \in T_{2j-1} \right] = p \left( 1 - p \right) + p^2 \left( 1 - p \right) + \cdots + p^{2j-2} \left( 1 - p \right) + p^{2j}.$$  

Thus we infer

$$\mathbb{E} \left( |T^*_{2j-1}| \right) = |T_{2j-1}| \cdot \Pr \left[ v \in T_{2j-1} \right]$$

$$\mathbb{E} \left( |T^*_{2j-1}| \right) = \frac{b-1}{b(1+p)} p \left( 1 - p \right) \left( 1 - p \right)^{\frac{b}{2j-1}} + p^{2j-1} \pm 1,$$

which completes the proof of Lemma 8.

Corollary 9. For $0 \leq i \leq \log_b n$, we have

$$\mathbb{E} \left( |T^*_i| \right) = \frac{b-1}{b(1+p)} p \left( 1 - p \right)^{\frac{b}{2j}} + p^{2j} \pm 1.$$  


Summing over all $i$ with $0 \leq i \leq \log_b n$, we have the following.

**Corollary 10.**

$$\mathbb{E}(|T_i^*|) = \sum_{i=0}^{\log_b n} \mathbb{E}(|T_i^*|) = \frac{b}{b + p}pn + O(\log n).$$

**Proof.** One may easily see that for $|x| \geq b \geq 2$, 

$$\sum_{i=0}^{\log_b n} \frac{1}{x^i} = \frac{x}{x-1} + O\left(\frac{1}{n}\right). \quad (6)$$

Corollary 9 yields that for $b \geq 2$

$$\sum_{i=0}^{\log_b n} \mathbb{E}(|T_j^*|) = \frac{b}{b - 1} \left[ O\left(\frac{1}{n}\right) + \frac{b}{b - 1} \mathbb{E}\left(\frac{1}{b}\right) + \frac{-b/p}{b - 1} p + O\left(\frac{1}{n}\right) \right] + O(\log n) \quad (7)$$

which completes the proof of Corollary 10. \qed

### 3.2. Concentration

Next we consider a concentration result of $|T_i^*|$. In other words, we show that $|T_i^*|$ is around its expectation with high probability. We will apply the following version of Chernoff bounds.

**Lemma 11** (Chernoff bound). Let $X_i$ be independent random variables such that $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$, and let $X = \sum_{i=1}^{n} X_i$. Then for any $\lambda \geq 0$,

$$\Pr[X \geq (1 + \lambda)\mathbb{E}(X)] \leq e^{-\frac{\lambda^2}{2}\mathbb{E}(X)}, \quad (8)$$

$$\Pr[X \leq (1 - \lambda)\mathbb{E}(X)] \leq e^{-\frac{\lambda^2}{2}\mathbb{E}(X)}. \quad (9)$$

In particular, for $0 \leq \lambda \leq 1$,

$$\Pr[|X - \mathbb{E}(X)| \geq \lambda\mathbb{E}(X)] \leq 2e^{-\frac{\lambda^2}{2}\mathbb{E}(X)}. \quad (10)$$

We first consider the case when $0 \leq i \leq 0.9 \log_b n$.

**Lemma 12.** For $0 \leq i \leq 0.9 \log_b n$, we have

$$|T_i^*| = \mathbb{E}(|T_i^*|) + O\left(\sqrt{pn \log \log n}\right) \quad (11)$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log \log n)^2}$. 


Proof. Fix $i$. If $k \in T_i \subset [n]$, then let

$$X_k = \begin{cases} 1 & \text{with probability } p^* \\ 0 & \text{with probability } 1 - p^* \end{cases}$$

where $p^* = \Pr[v \in T_i \text{ is in } T^*_i]$. Otherwise, let $X_k = 0$ with probability 1. Let $X = \sum_{k=1}^{n} X_k$. Observe that

$$X = |T^*_i|$$

as random variables.

Note that for each $k \in T_i$, the event that $k \in T^*_i$ depends only on the events that $v \in [n]$, where the vertices $v$ are in the component of $D$ containing $k$ and $v \leq k$. Hence, $X_k$ are independent for all $k \in T_i$. Therefore we are able to use Chernoff bounds (Lemma 11) for a concentration result of $X$.

Set $\lambda = \frac{\log \log n}{\sqrt{\mathbb{E}(X)}}$. Note that $0 \leq \lambda \leq 1$ for $0 \leq i \leq 0.9 \log_b n$ since

$$\mathbb{E}(X) \geq \Omega \left( \frac{pn \varepsilon_p}{b^i} \right) \geq \Omega \left( \frac{pn \varepsilon_p}{n^{0.9}} \right) = \Omega \left( \varepsilon_p n^{0.1} \right),$$

where $\varepsilon_p$ is a positive constant such that $\varepsilon_p \to 0$ as $p \to 1$. The inequality (10) yields that

$$\Pr[X = \mathbb{E}(X) + 2(\log \log n)^2] \leq e^{-\frac{1}{3}(\log \log n)^2}. \quad (13)$$

(14) Corollary 9 yields that $\mathbb{E}(|X|) = O(pn)$, and hence, we infer that

$$X = \mathbb{E}(X) + O(\sqrt{p_n} \log \log n)$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log \log n)^2}$. This together with (12) completes the proof of Lemma 12. $\square$

Next we consider the remaining case when $0.9 \log_b n \leq i \leq \log_b n$.

Lemma 13. For $0.9 \log_b n \leq i \leq \log_b n$, we have $|T^*_i| = O((pn)^{0.1})$, with probability at least $1 - e^{-O(\log \log n)^2}$.

Proof. We define a random variable $X$ as in (12), that is, $X = |T^*_i|$. Set $\lambda = \frac{2(\log \log n)^2}{\mathbb{E}(X)}$. The inequality (8) yields that

$$\Pr[X \geq (1 + \lambda)\mathbb{E}(X)] \leq e^{-\frac{1}{2} \mathbb{E}(X)} = e^{-O(\log \log n)^2},$$

and hence,

$$\Pr[ X \geq \mathbb{E}(X) + 2(\log \log n)^2] \leq e^{-O(\log \log n)^2}. \quad (14)$$

In other words,

$$X \leq \mathbb{E}(X) + 2(\log \log n)^2 \quad (15)$$

with probability at least $1 - e^{-O(\log \log n)^2}$. 7
Corollary 9 gives that
\[ E(X) = O \left( \frac{pn}{b^i} \right) = O \left( p^{0.1} \right) = O \left( (pn)^{0.1} \right), \quad (16) \]
where the second inequality holds for \( i \geq 0.9 \log_b n \). Thus, combining (15) and (16) completes the proof of Lemma 13. □

Now we are ready to show Lemma 5.

**Proof of Lemma 5.** We have that
\[ |T^*| = \sum_{i=1}^{\log_b n} |T^*_i| = \sum_{i=1}^{[0.9 \log_b n]} |T^*_i| + \sum_{i=[0.9 \log_b n]+1}^{\log_b n} |T^*_i|. \]

Lemmas 12 and 13 give that
\[ |T^*| = \sum_{i=1}^{\log_b n} E(|T^*_i|) + O(\sqrt{pm \log n \log \log n}), \]
with probability at least
\[ 1 - (\log_b n) \cdot 2e^{-\frac{1}{2}(\log \log n)^2} = 1 - 2e^{\log \log n - \frac{1}{2}(\log \log n)^2} = 1 - o(1). \quad (17) \]

This together with Corollary 10 implies that with high probability
\[ |T^*| = \frac{b}{b + p} pm + O(\sqrt{mn \log n \log \log n}), \]
which completes the proof of Lemma 5. □

**Acknowledgement.** The author thanks Yoshiharu Kohayakawa for his helpful comments and suggestions, and thanks Jaigyoung Choe for his support at Korea Institute for Advanced Study.

**References**

[1] S. Chowla. Solution of a problem of Erdős and Turan in additive-number theory. *Proc. Nat. Acad. Sci. India. Sect. A.*, 14:1–2, 1944.
[2] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. submitted, 70pp, 2010.
[3] P. Erdős. On a problem of Sidon in additive number theory and on some related problems. *Addendum, J. London Math. Soc.*, 19:208, 1944.
[4] P. Erdős and P. Turán. On a problem of Sidon in additive number theory, and on some related problems. *J. London Math. Soc.*, 16:212–215, 1941.
[5] M. Jimbo, M. Mishima, S. Janiszewski, A. Y. Teymorian, and V. D. Tonchev. On conflict-avoiding codes of length \( n = 4m \) for three active users. *IEEE Trans. Inform. Theory*, 53(8):2732–2742, 2007.
[6] Y. Kohayakawa, S. Lee, and V. Rödl. The maximum size of a Sidon set contained in a sparse random set of integers. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 159–171, Philadelphia, PA, 2011. SIAM.
Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. The number of Sidon sets and the maximum size of sidon sets contained in a sparse random set of integers. Accepted to Random Structures & Algorithms.

Y. Kohayakawa, T. Łuczak, and V. Rödl. Arithmetic progressions of length three in subsets of a random set. Acta Arith., 75(2):133–163, 1996.

J. Y.-T. Leung and W.-D. Wei. Maximal $k$-multiple-free sets of integers. Ars Combin., 38:113–117, 1994.

K. F. Roth. On certain sets of integers. J. London Math. Soc., 28:104–109, 1953.

M. Schacht. Extremal results for random discrete structures. submitted, 27pp, 2009.

E. Szemerédi. On sets of integers containing no $k$ elements in arithmetic progression. Acta Arith., 27:199–245, 1975. Collection of articles in memory of Jurii Vladimirovič Linnik.

D. Wakeham and D. Wood. On multiplicative Sidon sets. arXiv:1107.1073.

E. T. H. Wang. On double-free sets of integers. Ars Combin., 28:97–100, 1989.

School of Mathematics, Korea Institute for Advanced Study (KIAS), Seoul 130-722, South Korea

E-mail address: sjlee242@gmail.com