On asymptotic bases which have distinct subset sums

Sándor Z. Kiss [*], Vinh Hung Nguyen [†]

September 9, 2021

Abstract

Let k and l be positive integers satisfying $k \geq 2, l \geq 1$. We say a set $\mathcal{A}$ of positive integers is an asymptotic basis of order k if every large enough positive integer can be represented as the sum of k terms from $\mathcal{A}$. About 35 years ago, P. Erdős suggested a well-known question: Does there exist an asymptotic basis of order k where all the subset sums with at most l terms are pairwise distinct with the exception of finitely number of cases as long as $l \leq k-1$? In this paper, we prove the existence of an asymptotic basis of order $2k+1$ and all the sums of at most k elements of this asymptotic basis are pairwise different except for "small" numbers by using probabilistic tools.

2010 Mathematics Subject Classification: 11B13, 11B75.

Keywords and phrases: additive number theory, Sidon set, asymptotic basis, probabilistic method.

1 Introduction

Let $h, k \geq 2$ be integers. Let $\mathcal{A} \subset \mathbb{N}$ be an infinite set, where $\mathbb{N}$ denotes the set of all nonnegative integers. Now, for each $n \in \mathbb{N}$, we denote $r_k^*(\mathcal{A}, n)$ as

[*]Institute of Mathematics, Budapest University of Technology and Economics, H-1529 B.O. Box, Hungary; kisspest@cs.elte.hu; This author was supported by the National Research, Development and Innovation Office NKFIH Grant No. K115288 and K129335. This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Supported by the ÚNKP-19-4 New National Excellence Program of the Ministry for Innovation and Technology. Supported by the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

[†]Institute of Mathematics, Budapest University of Technology and Economics, H-1529 B.O. Box, Hungary; nguyenvinhhung108@gmail.com.

the number of solutions of the equation

$$a_1 + a_2 + \ldots + a_k = n,\ \ a_1 \in \mathcal{A}, \ldots, a_k \in \mathcal{A},\ \ a_1 \leq a_2 \leq \ldots \leq a_k.$$

We also define $r_k(\mathcal{A}, n)$ as the number of solutions of the equation

$$a_1 + a_2 + \ldots + a_k = n,\ \ a_1 \in \mathcal{A}, \ldots, a_k \in \mathcal{A},\ \ a_1 < a_2 < \ldots < a_k.$$

For a positive integer g, let $B_h[g] = \{\mathcal{A} \subset \mathbb{N} \mid r_h^*(\mathcal{A}, n) \leq g \text{ for every } n \in \mathbb{N}\}$. If $\mathcal{A} \in B_h[g]$, we say $\mathcal{A}$ is a $B_h[g]$ set and if $\mathcal{A} \in B_2[1]$, $\mathcal{A}$ is a Sidon set. Moreover, we say a set $\mathcal{A} \subset \mathbb{N}$ is an asymptotic basis of order k if there exists a positive number n_0 such that $r_k^*(\mathcal{A}, n) > 0$ for every $n > n_0$.

Over many years, the $B_h[g]$ sets which are asymptotic bases of some order were investigated by many authors. According to a famous conjecture of Erdős and Turán [8], an asymptotic basis of order 2 cannot be a $B_2[g]$ set. In [4] and [5], P. Erdős, A. Sárközy and V. T. Sós asked if there exists a Sidon set which is an asymptotic basis of order 3. J. M. Deshouillers and A. Plagne in [2] introduced a construction for a Sidon set which is an asymptotic basis of order at most 7. The existence of Sidon sets which are asymptotic bases of order 5 was also proved with the help of probabilistic tools in [10]. In addition, there is an improvement: in [1] and [12], it was showed that there exists Sidon sets that are asymptotic bases of order 4. Also in [1], it was proved the existence of $B_2[2]$ sets which are an asymptotic bases of order 3. In [6], Erdős asked for the largest possible value of the positive integer m such that all the possible 2^m subsums of a_i's are different, where $0 < a_1 < a_2 < ... < a_m \leq n$. Furthermore, in [3], P. Erdős raised a question which asks if there exists for every k an asymptotic basis of order k for which all the sums of the form $\sum_{i=1}^{l} \epsilon_i a_i$, where $\epsilon_i \in \{0, 1\}$ are all distinct except for a finite number of cases as long as $l \leq k - 1$.

In this paper, we will prove the existence of an asymptotic basis of order $2k + 1$ which satisfies the property that all $\sum_{i=1}^{k} \epsilon_i a_i$ with $\epsilon_i \in \{0, 1\}$ are pairwise distinct.

Namely, we prove the following theorem.

Theorem 1 *For every $k \geq 2$ integer, there exists a set which is an asymptotic basis of order $2k + 1$ and all the sums of the form $\sum_{i=1}^{k} \epsilon_i a_i$ with $\epsilon_i \in \{0, 1\}$ are all pairwise distinct.*

To prove Theorem 1, we apply the probabilistic method. In the next part, we give a short survey about it.

2 Probabilistic and combinatorial tools

To prove Theorem 1, we use the material of [13], which is based on the probabilistic method due to Erdős and Rényi. There is an excellent summary of this method in the book of Halberstam and Roth [9]. In this paper, we denote the probability of an event A by $P(A)$, and the expected value of a random variable X by $E(X)$. Let Ω denote the set of strictly increasing sequences of positive integers.

Lemma 1 *Let $\alpha_1, \alpha_2, \alpha_3, \dots$ be real numbers which satisfies*

$$0 \leq \alpha_n \leq 1 \ \ (n = 1, 2, 3, \dots).$$

Then there exists a probability space (Ω, X, P) with the following properties:

(i) For every natural number n, the event $\mathcal{E}^{(n)} = \{\mathcal{A}\colon\ \mathcal{A} \in \Omega,\ n \in \mathcal{A}\}$ is measurable, and $P(\mathcal{E}^{(n)}) = \alpha_n$.

(ii) The events $\mathcal{E}^{(1)}$, $\mathcal{E}^{(2)}$, $\dots$ are independent.

See Theorem 13. in [9], p. 142. We denote the indicator function of the event $\mathcal{E}^{(n)}$ by $\varrho(\mathcal{A}, n)$ i.e.,

$$\varrho(\mathcal{A}, n) = \begin{cases} 1, \text{ if } n \in \mathcal{A} \\ 0, \text{ if } n \notin \mathcal{A}. \end{cases}$$

We can easily see that for some $\mathcal{A} = \{a_1, a_2, ...\} \in \Omega$, we can calculate $r_h(\mathcal{A}, n)$, i.e., the number of solutions of $a_{i_1} + a_{i_2} + ... + a_{i_h} = n$ with $a_{i_1} \in \mathcal{A}$, $a_{i_2} \in \mathcal{A}$, ..., $a_{i_h} \in \mathcal{A}$ and $1 < a_{i_1} < \ldots < a_{i_h} < n$ by

$$r_h(\mathcal{A}, n) = \sum_{\substack{(a_{i_1}, a_{i_2}, \dots, a_{i_h}) \in \mathbb{N}^h \\ 1 \leq a_{i_1} < \ldots < a_{i_h} < n \\ a_{i_1} + a_{i_2} + \ldots + a_{i_h} = n}} \varrho(\mathcal{A}, a_{i_1}) \varrho(\mathcal{A}, a_{i_2}) \dots \varrho(\mathcal{A}, a_{i_h}).$$

We also need the following lemma for the proof of the theorem:

Lemma 2 *(Borel-Cantelli) Let $X_1, X_2, ...$ be a sequence of events in a probability space. If*

$$\sum_{j=1}^{+\infty} P(X_j) < \infty,$$

then with probability 1, at most a finite number of the events X_j can occur.

See [9], p. 135.
Another lemma due to Erdős and Tetali is needed for our proof:

Lemma 3 *Let $Y_1, Y_2, ...$ be a sequence of events in a probability space. If $\sum_i P(Y_i) \leq \mu$, and κ is a positive integer then*

$$\sum_{\substack{(Y_1,\ldots,Y_\kappa)\\ independent}} P(Y_1 \cap \ldots \cap Y_\kappa) \leq \frac{\mu^\kappa}{\kappa!}.$$

The proof of this lemma is provided in [7]. (We say the events $Y_1, Y_2, ..., Y_\omega$ are independent if for all subsets $I \subset \{1, 2, \ldots, \omega\}$, $P(\cap_{i\in I} Y_i = \prod_{i\in I} P(Y_i))$.

3 Proof of the theorem

Let $k \geq 2$ be fixed and $\alpha = \frac{2}{4k+1}$. Define the sequence α_n in Lemma 1 by

$$\alpha_n = \frac{1}{n^{1-\alpha}},$$

so that $P(\{\mathcal{A} \in \Omega,\ n \in \mathcal{A}\}) = \frac{1}{n^{1-\alpha}}$. It was already proved in [13] that $\mathcal{A}$ is an asymptotic basis of order $2k+1$, with probability 1. Therefore, to complete the proof, we need to show that, starting out from such an $\mathcal{A}$, we can construct an asymptotic basis of order $2k+1$, where all the sums

$$\sum_{i=1}^{k} \epsilon_i a_i$$

with $\epsilon_i = 0$ or $\epsilon_i = 1$ for every $i \geq 1$ are all distinct. To do that, we will prove that after deleting finitely many elements from $\mathcal{A}$, we will get a new set $\mathcal{C}$ such that $r_1(\mathcal{C}, n) + r_2(\mathcal{C}, n) + r_3(\mathcal{C}, n) + \ldots + r_k(\mathcal{C}, n) \leq 1$ with probability 1 for every $n \geq 1$, where $r_1(\mathcal{C}, n) = \varrho(\mathcal{C}, n)$. Furthermore, we will show that the above deletion does not destroy the asymptotic basis property.
Applying the following lemma with $w = 2k+1$ we get that $\mathcal{A}$ is an asymptotic basis of order $2k+1$, with probability 1.

Lemma 4 *Let $w \geq 2$ be a fixed integer and let $P(\{\mathcal{A} : \mathcal{A} \in \Omega, n \in \mathcal{A}\}) = \frac{1}{n^{1-\alpha}}$ where $\alpha > \frac{1}{w}$. Then with probability 1, $r_w(\mathcal{A}, n) > cn^{w\alpha-1}$ for every sufficiently large n, where $c = c(\alpha, w)$ is a positive constant.*

This is Lemma 3 in [13] and the proof can be found in [11]. It was also proved in [13] that deleting finitely many elements from $\mathcal{A}$ we get a set $\mathcal{B}$

which is a $B_k[1]$ set with probability 1. Now we show that removing finitely many elements from such a $\mathcal{B}$, we obtain a set $\mathcal{C}$ such that $r_1(\mathcal{C},n)+r_2(\mathcal{C},n)+r_3(\mathcal{C},n)+\ldots+r_k(\mathcal{C},n) \leq 1$ with probability 1. Since almost surely $\mathcal{B} \in B_k[1]$, then clearly $r_i^*(\mathcal{B},n) \leq 1$ for every $2 \leq i \leq k$ and $n \geq 1$, with probability 1, which obviously implies that $r_i(\mathcal{B},n) \leq 1$ for every $2 \leq i \leq k$ and $n \geq 1$, with probability 1.

In next step, we define the sets $\beta_j(n) = \{(a_1, a_2, \ldots, a_j) |\, a_1 < a_2 < \ldots < a_j,\ a_1 + a_2 + \ldots + a_j = n\}$ for $j = 1, 2, 3, \ldots, k$. Clearly, $\beta_1(n) = \{n\}$. With these notations, we define the set $\beta(n) = \cup_{j=1}^{k} \beta_j(n)$. For every $2 \leq i \leq j \leq k$, we say any two representations $(a_1, \ldots, a_i) \in \beta(n)$ and $(b_1, \ldots, b_j) \in \beta(n)$ are disjoint if $a_m \neq b_p$ for any $1 \leq m \leq i$ and $1 \leq p \leq j$. We can define $H(\beta(n)) = \{T \subseteq \beta(n) \mid \text{every two representations in } T \text{ are disjoint}\}$. Let $f_{\mathcal{A}}(n)$ and $f_{\mathcal{B}}(n)$ denote the size of the maximal collection of pairwise disjoint representations of n in $\mathcal{A}$ and $\mathcal{B}$ as the sum of at most k terms, respectively. In the next step, we prove that almost always there exists an n_1 such that

$$f_{\mathcal{B}}(n) \leq 1$$

for every $n \geq n_1$. To do this we need the following estimation for the expectation of $r_l(\mathcal{A},n)$.

Lemma 5 *For every $1 \leq l \leq k$, $E(r_l(\mathcal{A},n)) \leq n^{-1+l\alpha+o(1)}$ for every n*

The proof of Lemma 6 is similar to (5) in [10]. For the sake of completeness we present it. By $\frac{n}{l} < a_l$, we have:

$$E(r_l(\mathcal{A},n)) = \sum_{\substack{a_1+\ldots+a_l=n \\ 1\leq a_1<\ldots<a_l<n}} P(a_1 \in \mathcal{A}) \ldots P(a_l \in \mathcal{A})$$

$$= \sum_{\substack{a_1+\ldots+a_l=n \\ 1\leq a_1<\ldots<a_l<n}} \frac{1}{(a_1 \ldots a_l)^{1-\alpha}} \leq n^{-1+\alpha+o(1)} \sum_{\substack{a_1+\ldots+a_l=n \\ 1\leq a_1<\ldots<a_l<n}} \frac{1}{(a_1 \ldots a_{l-1})^{1-\alpha}}$$

$$\leq n^{-1+\alpha+o(1)} \sum_{\substack{1\leq a_i\leq n \\ i=1,2,\ldots,l-1 \\ 1\leq a_1<\ldots<a_{l-1}\leq n}} \frac{1}{(a_1 \ldots a_{l-1})^{1-\alpha}} \leq n^{-1+\alpha+o(1)} \sum_{1\leq a_1\leq n} (\frac{1}{a_1^{1-\alpha}})^{l-1}$$

$$= n^{-1+\alpha+o(1)} \big(n^{\alpha+o(1)}\big)^{l-1} = n^{-1+l\alpha+o(1)}.$$

It is clear that if we have two disjoint representations of n as the sum of at most k distinct terms S_1 and S_2, it implies the fact that two events $S_1 \subset \mathcal{A}$ and $S_2 \subset \mathcal{A}$ are independent. Using the fact that $\mathcal{B}$ is a subset of $\mathcal{A}$, $E(r_1(\mathcal{A},n)) = P(n \in \mathcal{A})$, Lemma 3 and Lemma 6, we have:

$$P(f_{\mathcal{B}}(n) \geq 2) \leq P(f_{\mathcal{A}}(n) \geq 2) \leq P(\cup_{\substack{T\in H(\beta(n)) \\ |T|=2}} \cap_{S\in T} S \subset \mathcal{A})$$

$$\leq \sum_{\substack{(S_1,S_2)\\ disjoint}} P((S_1 \subset \mathcal{A}) \cap (S_2 \subset \mathcal{A})) \leq \frac{E(f_{\mathcal{A}}(n))^2}{2!}$$

$$\leq \frac{1}{2}(E(r_1(\mathcal{A},n)) + E(r_2(\mathcal{A},n) + \ldots + r_k(\mathcal{A},n)))^2$$

$$\leq \frac{1}{2}(E(r_1(\mathcal{A},n)) + E(r_2(\mathcal{A},n)) + \ldots + E(r_k(\mathcal{A},n)))^2$$

$$\leq \frac{1}{2}(n^{-1+\alpha+o(1)} + \ldots + n^{-1+k\alpha+o(1)})^2 \leq \frac{k^2}{2} n^{-2+2k\alpha+o(1)}.$$

We can see that for every $k \geq 2$ we have

$$-2 + 2k\alpha = -2 + \frac{4k}{4k+1} = -1 - \frac{1}{4k+1} < -1.$$

By the Borel-Cantelli lemma, we can conclude: there almost surely exists $n_1 \geq 1$ such that $f_{\mathcal{B}}(n) \leq 1$ for every $n \geq n_1$.

Starting out from such a $\mathcal{B}$ we define a new set $\mathcal{C} = \mathcal{B} \setminus \mathcal{D}$ where $\mathcal{D} = \{a \in \mathcal{B} |\, a \leq n_1\}$. Thus we have $f_{\mathcal{C}}(n) \leq 1$ for every $n \geq 1$, where $f_{\mathcal{C}}(n)$ denotes the size of the maximal collection of pairwise disjoint representations of n as the sum of at most k terms from $\mathcal{C}$. Now we denote $F(\mathcal{C},n) = r_1(\mathcal{C},n) + r_2(\mathcal{C},n) + r_3(\mathcal{C},n) + \ldots + r_k(\mathcal{C},n)$. In the next step, we will prove that $F(\mathcal{C},n) \leq 1$ for every n. If $F(\mathcal{C},n) \geq 2$ then there must be at least two indices $1 \leq i \leq k$ and $1 \leq j \leq k$ where $i \neq j$ such that $r_i(\mathcal{C},n) = r_j(\mathcal{C},n) = 1$ because of the definition of the set $\mathcal{B}$ and the fact that $\mathcal{C} \subset \mathcal{B}$. Then there exists two representations $(a_1,\ldots,a_i) \subset \mathcal{C}$ and $(b_1,\ldots,b_j) \subset \mathcal{C}$ satisfying

$$a_1 + a_2 + \ldots + a_i = b_1 + b_2 + \ldots + b_j.$$

It cannot happen that for every $1 \leq m \leq i,\ 1 \leq q \leq j,\ a_m \neq b_q$, otherwise it would violate $f_{\mathcal{C}}(n) \leq 1$.

Based on this observation, it is obvious that there exists some $a_{m_1} = b_{q_1}, \ldots, a_{m_l} = b_{q_p}$ where $1 \leq m_1 < m_2 < \ldots < m_l \leq i$ and $1 \leq q_1 < q_2 < \ldots < q_p \leq j$. After cancelling the equal elements of both sides of the equation, it results in another equation

$$a_{m_1'} + a_{m_2'} + \ldots + a_{m_{l'}'} = b_{q_1'} + b_{q_2'} + \ldots + b_{q_{p'}'}, \tag{1}$$

where $1 \leq m_1' < m_2' < \ldots < m_{l'}' \leq i,\ 1 \leq q_1' < q_2' < \ldots < q_{p'}' \leq j$ and every element of the left-hand side and right-hand side of (1) is pairwise distinct, which is a contradiction again.

In the final step, we will show that $\mathcal{C}$ is an asymptotic basis of order $2k+1$. In other words, the removals of finitely many elements don't demolish the

asymptotic basis property of an asymptotic basis of the same order. We are now proving this statement by contradiction. Assume that there exists infinitely many positive integers N which cannot be expressed as the sum of $2k+1$ elements of $\mathcal{C}$. By our assumption, it follows that every representation of N as the sum of $2k+1$ terms from $\mathcal{A}$ must contain at least one term which comes from the finite set $\mathcal{A} \setminus \mathcal{C}$. Simultaneously, we have: there exists a positive number G such that $\mathcal{A}$ is a $B_{2k}[G]$ set; the proof of this statement can be found in [13], on page 6. According to Lemma 5, we can choose a large enough positive integer N such that $r_{2k+1}(\mathcal{A}, N) > cN^{\frac{1}{4k+1}}$. By the pigeon-hole principle, there must exist one number $x \in \mathcal{A} \setminus \mathcal{C}$ belonging to at least $\frac{r_{2k+1}(\mathcal{A},N)}{|\mathcal{A}\setminus\mathcal{C}|}$ representations of N as the sum of $2k+1$ terms from $\mathcal{A}$. Since $\mathcal{A} \in B_{2k}[G]$ as we stated above, it follows that

$$\frac{cN^{\frac{1}{4k+1}}}{|\mathcal{A} \setminus \mathcal{C}|} < \frac{r_{2k+1}(\mathcal{A}, N)}{|\mathcal{A} \setminus \mathcal{C}|} \leq r_{2k}(\mathcal{A}, N-x) \leq G,$$

which contradicts the large enough property of N. This completes the proof of Theorem 1.

References

[1] J. Cilleruelo. *On Sidon sets and asymptotic bases*, Proc. Lond. Math. Soc., **111**, (2015) 1206-1230.

[2] J. M. Deshouillers, A. Plagne. *A Sidon basis*, Acta Mathematica Hungarica, **123**, (2009) 233-238.

[3] P. Erdős. *Some applications of probability methods to number theory*, Mathematical statistics and applications, Vol. B (Bad Tatzmannsdord, 1983), (1985) 1-18.

[4] P. Erdős, A. Sárközy, V. T. Sós. *On additive properties of general sequences*, Discrete Mathematics, **136**, (1994) 75-99.

[5] P. Erdős, A. Sárközy, V. T. Sós. *On sum sets of Sidon sets I.*, Journal of Number Theory, **47**, (1994) 329-347.

[6] P. Erdős, J. Spencer. *Probabilistic Methods in Combinatorics*, Akadémiai Kiadó, Budapest, 1974.

[7] P. Erdős, P. Tetali. *Representations of integers as the sum of k terms*, Random Structures and Algorithms, **1**, (1990) 245-261.

[8] P. Erdős, P. Turán. *On a problem of Sidon in additive number theory, and some related problems*, J. London Math. Soc., **16**, (1941) 212-215.

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

[10] S. Z. Kiss. *On Sidon sets which are asymptotic bases*, Acta Mathematica Hungarica, **128**, (2010) 46-58.

[11] S. Z. Kiss. *On generalized Sidon sets which are asymptotic bases*, Annales Univ. Sci. Budapest. Eötvös, **57**, (2014) 149-160.

[12] S. Z. Kiss, E. Rozgonyi, Cs. Sándor. *On Sidon sets which are asymptotic bases of order 4*, Functiones et Approximatio Comm. Math., **51**, (2014) 393-413.

[13] S. Z. Kiss, C. Sándor. *Generalized asymptotic Sidon basis*, Discrete Mathematics, **344**, (2021) 112208.