HIGH ORDER COMPLEMENTARY BASES OF PRIMES

V. H. Vu [1]
Department of Mathematics, University of California at San Diego
La Jolla, CA 92093, USA
vanvu@ucsd.edu

Received: 3/24/02, Accepted: 10/22/02, Published: 10/23/02

Abstract

We show that there is a set $X \subset \mathbf{N}$ with density $O(\log n)$ such that every sufficiently large natural number can be represented as sum of two elements from X and a prime. The density is a $\log^{1/2} n$ factor off being best possible.

1. Introduction

A central notion in additive number theory is that of bases. A subset X of $\mathbf{N}$ is a basis of order k if every sufficiently large number $n \in \mathbf{N}$ can be represented as a sum of k elements of X. Here and later $\mathbf{N}$ denotes the set of natural numbers.

In this note, we are working with a related notion of complementary bases. Given a set $A \subset \mathbf{N}$, a set $X \subset \mathbf{N}$ is a complementary basis of order k of A if every sufficiently large natural number can be written as a sum of an element in A and k elements in X. All asymptotic notations are used under the assumption that $n \to \infty$. The logarithms have natural base.

Consider the set P of primes. Since P has density $n/\log n$, it is clear by the pigeon hole principle that a complementary basis of order k of P should have density $\Omega(\log^{1/k} n)$. As far as we know, it is still an open question to determine, even for the case $k = 1$, that whether there is a complementary basis of P with density $O(\log^{1/k} n)$. In [1], Erdős shown that for $k = 1$, there is a complementary base of density $O(\log^2 n)$. In a recent paper, Ruzsa [5], improving a result of Kolountzakis [4], showed that there exists a set X of density $\omega(n) \log n$, where $\omega(n)$ is a function tending to infinity arbitrarily slow in n, such that the set $X + P$ has density one (i.e., almost all natural numbers can be represented as a sum of an element from X and a prime).

[1] Research supported in part by grant RB091G-VU from UCSD, by NSF grant DMS-0200357, and by a Sloan fellowship.

AMS Subject Classificiation Numbers: 05, 11.
Key words: Complementary bases, the probabilistic method, primes.

In this note, we focus on the case $k = 2$. Our main theorem is

Theorem 1.1 *P has a complementary basis of order 2 and density $O(\log n)$.*

Corollary 1.2 *For all $k \geq 2$, P has a complementary basis of order k and density $O(\log n)$.*

We conjecture that

Conjecture. *For any fixed k, P has a complementary basis of order k and density $O(\log^{1+o(1))/k} n)$.*

The probabilistic method is the only effective approach so far to this type of problems. It seems that to prove even the density $O(\log^{2/k} n)$ for $k \geq 3$ requires a new tool from probability theory. Such a tool should surely be of independent interest. The first step towards the conjecture might be to prove a bound with the exponent decreasing in k.

2. Proof of Theorem 1.1

We construct the claimed complementary basis X by the random method. For each $x \in \mathbf{N}$, choose x to be in X with probability $p_x = \mathbf{c}/x$, where $\mathbf{c}$ is a positive constant to be determined. Let t_x be the binary random variable representing the choice of x (thus $t_x = 1$ with probability p_x and 0 with probability $1 - p_x$). We skip the fairly easy proof of the fact that almost surely, $X(m) = O(\log m)$ for every m, i.e, X has the right density; the interested readers might want to consider this as a warm-up exercise. Now let us consider the number representations of n as sum of a prime and two elements from X. This number is a random variable depending on the t_j's, $j < n$ and can be expressed as follows

$$Y_n = \sum_{p<n} \sum_{i+j=n-p} t_i t_j,$$

where in the second sum we do not count permutations. Here and later in a sum over p we understand that p is a prime. We next show that there is a constant n_0 such that with probability at least 1/2, $Y_n \geq 1$ for all $n \geq n_0$. To achieve this goal, it suffices to prove that for all sufficiently large n, $Pr(Y_n = 0) \leq n^{-2}$ (notice that the sum $\sum_{n=n_0}^{\infty} n^{-2}$ goes to 0 as n_0 tends to infinity). It has turned out that it is much more convenient to work with the following truncation of Y_n

$$Y'_n = \sum_{p \leq n-2n^{2/3}} \sum_{i+j=n-p;\ i,j \geq n^{2/3}} t_i t_j.$$

In the following, we shall prove

$$Pr(Y' = 0) \leq n^{-2}. \tag{1}$$

Central to the proof of (1) is the following

Lemma 2.3 *For all sufficient large n*

$$\sum_{p \leq n - n^{2/3}} \frac{1}{n-p} = \Theta(1),$$

and

$$\sum_{p \leq n - 2n^{2/3}} \frac{1}{n-p} = \Theta(1).$$

Proof of Lemma 2.3. To verify the first equality, let us set $n_1 = n - n^{2/3}$ and $n_l = n_{l-1} - n_{l-1}^{2/3}$ for all $l = 2, 3, 4, \ldots, s$ where s is the first place where $n_s \leq n/2$. It is a routine to show by induction that

$$n - \frac{ln^{2/3}}{2} \geq n_l \geq n - ln^{2/3}. \tag{2}$$

Let P_l denote the set of primes in the interval $[n_l, n_{l-1})$. It is clear, by (2), that for all $p \in P_l$,

$$\frac{2}{ln^{2/3}} \geq \frac{1}{n-p} \geq \frac{1}{ln^{2/3}}. \tag{3}$$

On the other hand, it is a well-known fact in number theory that the number of primes in the interval $[m - m^{2/3}, m)$ is $\Theta(m^{2/3}/\log m)$ for every sufficiently large m (see, for instance, [2]). Thus

$$|P_l| = \Theta(n_{l-1}^{2/3}/\log n_{l-1}) = \Theta(n^{2/3}/\log n)$$

for all l. This and (3) yield

$$\sum_{l=2}^{s} \sum_{p \in P_l} \frac{1}{n-p} = \Theta(\log^{-1} n \sum_{l=2}^{s} 1/l) = \Theta(\log^{-1} n \times \log s) = \Theta(1). \tag{4}$$

To complete the proof of the first equality, notice that $\sum_{p\geq n/2} 1/(n-p) \leq \sum_{j=n/2}^{n} 1/j = O(\frac{n}{\log n}\frac{1}{n}) = o(1)$. The second equality follows easily. $\mathcal{Q}.\mathcal{E}.\mathcal{D}$

The second main ingredient of our proof is a large deviation bound, due to Janson [3]. Let $z_1, \ldots, z_m$ be independent indicator random variables and consider a random variable $Y = \sum_\alpha I_\alpha$ where each I_α is the product of few z_j's. We write $\alpha \sim \beta$ if there is some z_j which occurs in both I_α and I_β. Furthermore, set $\Delta = \sum_{\alpha\sim\beta} \mathbf{E}(I_\alpha I_\beta)$. To this end, $\mathbf{E}(A)$ denotes the expectation of the random variable A.

Theorem 2.4 *(Janson) For any Y as above and any positive number ε*

$$Pr(Y \leq (1-\varepsilon)\mathbf{E}(Y)) \leq e^{-\frac{(\varepsilon \mathbf{E}(Y))^2}{2(\mathbf{E}(Y)+\Delta)}}.$$

From Theorem 2.4, one can routinely derive the following lemma. For a pair $1 \leq i, j \leq m$, we write $i \sim j$ if there is some α such that I_α contains both z_i and z_j.

Lemma 2.5 *There is a positive constant r such that the following holds. If each term I_α in Y is the product of exactly 2 random variables and for all $m \geq i \geq 1$, $\mathbf{E}(Y) \geq r \log n\Big(\mathbf{E}(\sum_{j\sim i} t_j) + 1\Big)$, then*

$$Pr(Y = 0) \leq n^{-2}.$$

We now apply Lemma 2.5 to Y'_n. Let us notice that in our setting, $i \sim j$ if and only if there is a prime number p such that $i + j + p = n$. Therefore,

$$\mathbf{E}(\sum_{j\sim i} t_j) = \sum_{p\leq n-i-n^{2/3}} \frac{\mathbf{c}}{n-i-p} \leq \sum_{p\leq m-m^{2/3}} \frac{\mathbf{c}}{m-p},$$

where $m = n - i$. By Lemma 2.3, there is a constant a such that $\mathbf{E}(\sum_{j\sim i} t_j) \leq a\mathbf{c}$. Moreover, a simple calculation yields that

$$\mathbf{E}(\sum_{i+j=n-p, i,j\geq n^{2/3}} t_i t_j) = \Omega(\mathbf{c}^2 \frac{\log(n-p)}{n-p}).$$

This and Lemma 2.3 together imply

$$\mathbf{E}(Y') = \Omega(\sum_{p\leq n-2n^{2/3}} \frac{\log(n-p)}{n-p}) = \Omega(\mathbf{c}^2 \log n \sum_{p\leq n-2n^{2/3}} \frac{1}{n-p}) = \Omega(\mathbf{c}^2 \log n). \qquad (5)$$

(5) guarantees that there is a positive constant b such that $\mathbf{E}(Y') \geq b\mathbf{c}^2 \log n$. Thus, by increasing $\mathbf{c}$, we can guarantee that the condition of Lemma 2.5 is met and this completes the proof. $\mathcal{Q.E.D}$

Acknowledgement. The author would like to thank Andrew Granville for a correction concerning the result of Ruzsa quoted in the introduction.

References

[1] P. Erdős, Problems and results in additive number theory, *Colloque sur la Théory de Nombre (CBRM)* (Bruxelles, 1956), 127-137

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

[3] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms,* 1 (1990), 221-230.

[4] M. Kolountzakis, On the additive complements of the primes and sets of similar growth, *Acta Arith.* 77 (1996), no. 1, 1–8.

[5] I. Ruzsa, On the additive completion of primes. *Acta Arith. 86* (1998), no. 3, 269–275.