ACTA ARITHMETICA
LXXXVI.3 (1998)

On the additive completion of primes

by

Imre Z. Ruzsa (Budapest)

1. Introduction. We shall investigate sets B of positive integers with the property that, in various senses, most positive integers are contained in the sumset

$$S = \{p + b : b \in B,\ p \text{ prime}\}. \tag{1.1}$$

We are interested in how thin this set can be in terms of the counting function

$$B(x) = |B \cap [1, x]|.$$

Erdős [2] (see also [5]) proved the existence of a set B such that S contains all but finitely many natural numbers and

$$B(x) = O((\log x)^2).$$

Improving a result of Wolke [7], Kolountzakis [6] proved the existence of a set B such that $d(S) = 1$, where d denotes asymptotic density, and

$$B(x) = O(\log x \log\log x).$$

(It should be noted here that Wolke's completion, while slightly denser, has the interesting additional property of consisting exclusively of primes and neighbours of primes.)

We improve Kolountzakis' result as follows.

Theorem 1. (a) *For every* $\varepsilon > 0$ *there is a set* B *such that*

$$B(x) = O(\log x)$$

and the set S *defined in* (1.1) *satisfies* $\underline{d}(S) > 1 - \varepsilon$, *where* $\underline{d}$ *denotes lower asymptotic density.*

1991 *Mathematics Subject Classification*: 11P32, 11B75.

Supported by Hungarian National Foundation for Scientific Research, Grant No. T 017433.

(b) *For every function* $\omega(x) \to \infty$ *there exists a set* B *such that*

$$B(x) = O(\omega(x) \log x)$$

and the set S *defined in* (1.1) *satisfies* $d(S) = 1$.

As far as I know, no lower estimate has been published for $B(x)$ in either context. An obvious counting argument yields that if $d(S) = 1$, then

$$\liminf B(x)/\log x \geq 1.$$

I think that Theorem 1 above is best possible in the following sense.

Conjecture 1. *If* $d(S) = 1$, *then necessarily*

$$B(x)/\log x \to \infty.$$

I am unable to prove even the following weaker conjectures.

Conjecture 2. *If* S *contains all but finitely many natural numbers*, *then necessarily*

$$B(x)/\log x \to \infty.$$

Conjecture 3. *If* $d(S) = 1$, *then necessarily*

$$\limsup B(x)/\log x > 1.$$

In this direction we obtain the following result.

Theorem 2. *If* S *satisfies*

$$x - S(x) \leq x^{1 - \log\log\log x/\log\log x}$$

for large x (*hence a fortiori if* S *contains all but finitely many natural numbers*), *then*

$$\liminf B(x)/\log x \geq e^{\gamma},$$

where γ *denotes the Euler–Mascheroni constant.*

2. Proof of Theorem 1.

We use P to denote the set of primes, and the traditional $\pi(x)$ (rather than $P(x)$) for its counting function.

Let $0 < c_0 < 1$ be a constant with the following property:

$$\pi(x+y) - \pi(x) \sim y/\log x \tag{2.1}$$

holds uniformly in the range $x^{c_0} \leq y \leq x$ as $x \to \infty$ (2/3 is known to be such a number), and fix another constant c_1 with $c_0 < c_1 < 1$.

Both parts of the theorem will follow from the following finite version.

Lemma 2.1. *For every* $\varepsilon > 0$ *there is a* $K = K(\varepsilon)$ *and an* $N_0 = N_0(\varepsilon)$ *such that for* $N > N_0$ *we can always find a set* $B \subset [N^{c_0}, 2N^{c_0}]$ *of integers such that*

$$|B| \leq K \log N \tag{2.2}$$

and the set $S = P + B$ *satisfies*

$$S(x) \geq (1-\varepsilon)x \quad \text{for all } N^{c_1} \leq x \leq N. \tag{2.3}$$

Proof. We show the existence of such a set by a probabilistic argument. Write

$$I = [N^{c_0}, 2N^{c_0}] \cap \mathbb{N}, \qquad L = |I| = N^{c_0} + O(1).$$

Let B be a random subset of I such that each $n \in I$ is included into B independently with a common probability

$$\varrho = \frac{K \log N}{2L}$$

so that the expected number of elements is

$$\mathbf{E}(|B|) = \varrho L = \frac{K}{2} \log N.$$

We will show that with a proper choice of K we have

$$\mathbf{P}((2.3) \text{ holds}) > 1/N, \tag{2.4}$$

while

$$\mathbf{P}(|B| > K \log N) < 1/N, \tag{2.5}$$

so that there will be a set that satisfies both (2.3) and $|B| \leq K \log N$.

Note that we will not establish that most sets B with about $K \log N$ elements have property (2.3). We are able to do this with a more complicated argument than the one in the sequel, which also exploits further properties of the primes outside (2.1).

We show that (2.3) follows from the following related property of a set B: for each $x_j = N/2^j$ satisfying $N^{c_1} \leq x_j \leq N$ we have

$$S(x_j) > (1-\varepsilon/2)x_j. \tag{2.6}$$

To see how (2.6) implies (2.3), observe that the function $T(x) = x - S(x)$ is increasing. Hence for a general x, with, say, $x_j \geq x > x_{j+1} = x_j/2$, we find

$$T(x) \leq T(x_j) \leq \varepsilon x_j/2 < \varepsilon x$$

as wanted.

Now we estimate the probability that (2.6) holds for a fixed value of j. A given number $n \in [2N^{c_0}, N]$ will be in S if at least one number $n-p$, p prime, is in B. This is possible if $n - p \in I$, or $p \in n - I$. The number of such primes is

$$z_n = \pi(n - N^{c_0}) - \pi(n - 2N^{c_0}) \sim L/\log n$$

by property (2.1). We have $n \notin S$ only if none of these z_n integers is in B; the probability of this event is

$$\mathbf{P}(n \notin S) = (1-\varrho)^{z_n} \leq \exp(-\varrho z_n).$$

We have

$$\varrho z_n \sim \frac{K \log N}{2L} \cdot \frac{L}{\log n} = \frac{K \log N}{2 \log n} \geq K/2,$$

thus $\varrho z_n > K/3$ for large N uniformly in the above range of n. Hence

$$\mathbf{P}(n \notin S) < e^{-K/3}. \tag{2.7}$$

Consequently, the expectation of $T(x_j)$ satisfies

$$\mathbf{E}(T(x_j)) < e^{-K/3} x_j + 2N^{c_0}, \tag{2.8}$$

where the last term comes from the fact that (2.7) need not hold for $n < 2N^{c_0}$. Since our numbers satisfy $x_j \geq N^{c_1}$, the second summand of (2.8) is of a smaller order of magnitude than the first and we have

$$\mathbf{E}(T(x_j)) < 2e^{-K/3} x_j$$

for large N. From Markov's inequality we can now infer

$$\mathbf{P}(T(x_j) \geq (\varepsilon/2)x_j) \leq 4e^{-K/3}/\varepsilon \leq 1/2,$$

or in other words

$$\mathbf{P}(S(x_j) > (1-\varepsilon/2)x_j) \geq 1/2, \tag{2.9}$$

provided

$$e^{-K/3} \leq \varepsilon/8. \tag{2.10}$$

The property that (2.6) holds for a fixed value of j defines an increasing family of sets B (if a set B has this property, then so does every set containing B). Such increasing families are always positively correlated, that is, the probability that a random set is in the intersection of several such families is at least as high as the product of the corresponding probabilities. This is stated in this form in Alon and Spencer [1, Theorem 3.2] and is a corollary to either Fortuin, Kasteleyn and Ginibre's inequality [4] (the usual approach), or to an earlier inequality of Esary, Proschan and Walkup [3]. In our case this means that

$$\mathbf{P}(S(x_j) > (1-\varepsilon/2)x_j \text{ for all } j) \geq \prod_j \mathbf{P}(S(x_j) > (1-\varepsilon/2)x_j) \geq 2^{-J},$$

where J is the number of the values x_j considered. Since these are defined by $x_j = N/2^j > N^{c_1}$, we have $2^J \leq 2N^{1-c_1} < N$, which concludes the proof that (2.6), and hence (2.3) as well, holds at least with probability $1/N$.

To show (2.5) note that

$$\mathbf{E}(e^{|B|}) = (1-\varrho+\varrho e)^L \leq \exp(\varrho(e-1)L) = \exp\left(\frac{e-1}{2} K \log N\right).$$

Thus from Markov's inequality we get

$$\begin{aligned}\mathbf{P}(|B| > K\log N) &= \mathbf{P}(e^{|B|} > e^{K\log N})\\ &\le \exp\left(\frac{e-1}{2}K\log N - K\log N\right) = N^{K(e-3)/2},\end{aligned}$$

which is smaller than $1/N$ provided

$$K > 2/(3-e). \tag{2.11}$$

(Alternatively, we could refer to various forms of Bernstein's or Chernoff's inequality.)

We have established (2.4) and (2.5), thus the proof of Lemma 2.1 is complete. For K we obtain the value

$$K = 3\log(8/\varepsilon), \tag{2.12}$$

which satisfies (2.11) if $\varepsilon < 3/4$.

Proof of Theorem 1. (a) We define a sequence N_i starting from the N_0 of the lemma by the recursion

$$N_{i+1} = [N_i^{1/c_1}].$$

We apply the lemma to each $N = N_i$ to get a set $B_i \subset [N_i^{c_0}, 2N_i^{c_0}]$ satisfying $|B_i| \le K\log N_i$ and (2.3) for the set $S_i = P + B_i$ in the range $N_i^{c_1} \le x \le N_i$. We put $B = \bigcup B_i$. For the set $S = P + B$ we have evidently

$$S(x) \ge S_i(x) \ge (1-\varepsilon)x$$

in the above range, and since N_i is selected to make these overlap, we have this everywhere. To estimate $B(x)$ define k as the smallest subscript with $N_k^{c_0} > x$. This means that $N_{k-1} \le x^{1/c_0}$, thus

$$N_k \le N_{k-1}^{1/c_1} \le x^{1/(c_0c_1)}.$$

Observe that $\log N_i$ grows exponentially, hence

$$B(x) \le \sum_{i\le k} |B_i| = O(|B_k|) = O(\log N_k) = O(\log x)$$

as claimed.

(b) In the proof of part (a) we replace the constant ε by a decreasing sequence ε_i tending to 0 sufficiently slowly.

3. Proof of Theorem 2.

We prove Theorem 2 in the following form. Suppose that $\alpha < e^\gamma$ and $B \subset [1,x]$ satisfies $k = |B| \le \alpha\log x$. Then for some $c < 1$ and $x > x_0(\alpha)$ there are at least

$$x^{1-c\log\log\log x/\log\log x}$$

numbers up to x not of the form $p + b$, $b \in B$, p prime.

Let p_i denote the ith prime. Put

$$r = \left[\frac{\log x}{(\log\log x)^3} \right]$$

and $m = p_1 \dots p_r$. By an averaging argument we find a residue class $a \pmod{m}$ such that

$$l = \#\{b \in B : (a-b,m)=1\} \le \frac{\phi(m)}{m} k.$$

Let $b_1, \dots, b_l$ be the elements of B such that $(b-a,m)=1$. Let π be any permutation of the set $\{1,\dots,l\}$. We shall consider the integers n satisfying

$$(3.1) \qquad n \equiv a \pmod{m}, \qquad n \equiv b_j \pmod{p_{r+\pi(j)}}, \qquad j=1,\dots,l.$$

These integers occupy a residue class modulo M, where

$$M = m \prod_{j=1}^{l} p_{r+j} = \prod_{i=1}^{r+l} p_i.$$

We show that the majority of these numbers lies outside $P+B$. Indeed, suppose that $n = p+b$, or $p = n-b$. If b is one of $b_1,\dots,b_l$, say b_j, then $n-b$ is divisible by $p_{r+\pi(j)}$. If b is none of these, then

$$(n-b,m) = (a-b,m) > 1,$$

thus $n-b$ is divisible by one of $p_1,\dots,p_r$. Consequently, the prime p can only be one of $p_1,\dots,p_{r+l}$, and the total number of such integers n (for all permutations together) is at most $k(r+l)$.

To estimate the number of integers n satisfying (3.1) we first find a bound for M. By Mertens's theorem we have

$$\frac{\phi(m)}{m} = \prod_{j=1}^{r} \left(1 - \frac{1}{p_j}\right) \sim e^{-\gamma}(\log r)^{-1} \sim e^{-\gamma}(\log\log x)^{-1}.$$

Consequently,

$$l \le \frac{\phi(m)}{m} k \le (1+o(1))\alpha e^{-\gamma} \frac{\log x}{\log\log x}.$$

Write $L = l + r$. By the above estimate and the definition of r we have

$$L \le \alpha_1 \frac{\log x}{\log\log x}$$

for large x with any $\alpha_1 > \alpha e^{-\gamma}$. Since $\alpha < e^{\gamma}$, we can achieve this with $\alpha_1 < 1$.

Using the prime number theorem in the form $p_i \sim i \log i$ we obtain

$$M = \prod_{i=1}^{L} p_i < L!(\log L)^L e^{o(L)}.$$

By (3.2) this yields $M = o(x)$, in particular, $M < x$ for large x.

Each congruence (3.1) defines a residue class modulo M, thus at least $[x/M] \ge x/(2M)$ integers up to x. Since there are $l!$ possible choices for the permutation π, this means altogether at least

$$\frac{1}{2}\frac{l!}{M}x$$

numbers. As said above, of these numbers at most Lk can be in $P+B$, thus the number elements outside $P+B$ is at least

$$\frac{1}{2}\frac{l!}{M}x - Lk. \tag{3.2}$$

To estimate this number from below observe that

$$\frac{M}{l!} \le \frac{L!}{l!}(\log L)^L e^{o(L)} < L^r (\log L)^L e^{o(L)}$$
$$< \exp\left((\alpha_1 + o(1))\frac{\log x \log\log\log x}{\log\log x}\right).$$

Hence the first term of (3.2) is larger than $x^{1-c\log\log\log x/\log\log x}$ with any $c > \alpha_1$ for large x. Since $Lk = o((\log x)^2)$, the second term is of smaller order of magnitude than the first, and our claim is established for arbitrary $\alpha_1 < c < 1$.

References

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, New York, 1992.

[2] P. Erdős, *Some results on additive number theory*, Proc. Amer. Math. Soc. 5 (1954), 847–853.

[3] J. D. Esary, F. Proschan and D. W. Walkup, *Association of random variables, with applications*, Ann. Math. Statist. 38 (1967), 1366–1374.

[4] C. M. Fortuin and P. W. Kasteleyn, *Correlation inequalities on some partially ordered sets*, Comm. Math. Phys. 22 (1971), 89–103.

[5] H. Halberstam and K. F. Roth, *Sequences*, Clarendon, London, 1966 (2nd ed., Springer, New York, 1983).

[6] M. N. Kolountzakis, *On the additive complements of the primes and sets of similar growth*, Acta Arith. 77 (1996), 1–8.

[7] D. Wolke, *On a problem of Erdős in additive number theory*, J. Number Theory 59 (1996), 209–213.

Mathematical Institute of the Hungarian Academy of Sciences
Pf. 127, H-1364 Budapest, Hungary
E-mail: ruzsa@math-inst.hu

Received on 3.3.1998
and in revised form on 25.5.1998 (3348)