Acta Arith., 77 (1996), 1, 1-8

# On the Additive Complements of the Primes and Sets of Similar Growth

Mihail N. Kolountzakis*

August 1995

**Abstract**

1. We prove that there is a set of nonnegative integers $A$, with counting function $A(x) = \#(A \cap [1,x])$ satisfying

$$A(x) \le C \log x \log\log x, \quad C \text{ a positive constant,}$$

such that $\mathbb{N} = (A + \mathbb{P}) \cup E$, with $\mathbb{N}$ and $\mathbb{P}$ denoting the positive integers and the primes respectively and with the set $E$ having upper density 0. That is, $A$ is "almost" an additive complement of the primes and at the same time it has few elements (only $O(\log\log x)$ more than the a priori minimum). This improves and simplifies a recent result of Wolke.

2. We prove the existence of sets that are "hard" to complement. For example, we show that there is a set $A$, with $A(x) \sim x/\log x$, which cannot be complemented with any set $B$, that satisfies $\liminf_{x\to\infty} B(x)/\log^2 x < 1$. In other words $\mathbb{N} = A + B$ implies $B(x) \gtrsim \log^2 x$.

## 1 Introduction and Results

In 1954 Erdős [2], [6, Ch. 3] proved that there is a set of nonnegative integers $A \subseteq \mathbb{N}$, with counting function bounded by $C \log^2 x$, $C$ an absolute positive constant, which is an *additive complement* of the primes $\mathbb{P}$. That is, every positive integer $n$ can be written, in at least one way, as a sum $a + p$, with $a \in A$ and $p \in \mathbb{P}$ (otherwise written as $\mathbb{N} = A + \mathbb{P}$) and, at the same time, $A(x) = \#(A \cap [1,x]) \le C \log^2 x$. The proof of this fact was probabilistic. It was shown that almost all random sets $A$, drawn from the proper distribution, have the properties mentioned above.

This result has not been improved. Recently[1] Wolke [7] proved the following.

*For any function $h(x) \ge 0$ that tends to infinity as $x \to \infty$ there is a set $A \subseteq \mathbb{N}$ with*

$$A(x) \le C h(x) \log x \log\log x, \quad C \textit{ a positive constant,}$$

*School of Mathematics, Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540. E-mail address: `kolount@math.ias.edu`. Supported by NSF grant DMS 9304580.

[1]Announced at the International Conference on Analytic Number Theory, Monticello, Illinois, May 1995.

*and such that* $\mathbb{N} = (A+\mathbb{P}) \cup E$, *with* $E(x) = o(x)$, *as* $x \to \infty$.

We call a set $A$ as above an *almost complement of the primes* and $E$ the *exceptional* set corresponding to $A$.

In the first part of this note we improve the result of WOLKE. We give a probabilistic proof, simpler than that in [7], that there is an almost complement $A$ of the primes with

$$A(x) \le C \log x \log\log x,$$

with $C$ an absolute positive constant.

Note that the only obvious lower bound for the counting function of a complement $A$ of the primes is $A(x) \gtrsim \log x$, since, by the Prime Number Theorem, $\mathbb{P}(x) \sim x/\log x$. In the second part of this note we show that there are *prime-like* sets $A$ (i.e. sets with the growth of the primes) whose additive complements $B$ necessarily have counting functions satisfying

$$\liminf_{x\to\infty} \frac{B(x)}{\log^2 x} \ge 1.$$

That is, if ERDŐS's result is to be improved then one must use some properties of the primes besides their growth (the proof of ERDŐS's theorem is not using any such properties). The proof is again based on probability.

### 1.1 Notation

We write $A(x) = \#(A \cap [1,x])$ for the counting function of the sequence $A$. We also write $A^{\le x} = A \cap [1,x]$. By $a \ll b$ we mean $a \le Cb$, where $C$ is an absolute constant. The fact $\lim(a/b) = 1$, as an implicit quantity tends to a limit, will be denoted by $a \sim b$. The notation $a \gtrsim b$ means $\liminf(a/b) \ge 1$. The set of primes is denoted by $\mathbb{P}$. The letter $C$ is used for an absolute constant, not necessarily the same in all its occurences, even in the same equation. We say that the set $A$ *complements* the set $B$ if every sufficiently large integer can be written in at least one way as $a+b$, with $a \in A$ and $b \in B$.

We recall the following definition.

**Definition 1** *The upper density* $\rho(E)$ *of a set* $E \subseteq \mathbb{N}$ *is defined as*

$$\rho(E) = \limsup_{N\to\infty} \frac{\#(E\cap[1,N])}{N}.$$

A crucial observation here is that the set $E \subseteq \mathbb{N}$ has upper density 0 if and only if

$$\limsup_{N\to\infty} 2^{-N}\#(E\cap[1,2^N]) = 0.$$

### 1.2 New results

We state our results.

**Theorem 1** (Almost Additive Complement for the Primes)
*There is a set $A \subseteq \mathbb{N}$ with*

$$A(x) \sim C \log x \log\log x$$

*such that every integer $n$, not in an exceptional set $E$ of upper density $0$, can be represented at least once as*

$$n = a + p, \quad a \in A,\ p \text{ a prime.}$$

**Remark:** In Theorem 1, and for any $M > 0$, the counting function of the exceptional set $E$ can be taken to satisfy

$$E(x) \ll \frac{x}{\log^M x}.$$

For the purposes of Theorem 2 and Corollary 1 below, a set $B$ complements a set $A$ if all *sufficiently large* integers are representable as a sum of an element of $A$ and an element of $B$.

**Theorem 2** (Sets that are hard to complement)
*Let the function $\phi(x) \ge 1$ increase to infinity with $x$ and assume that the function $\psi(x) \ge 1$ satifies $\lim_{x\to\infty} \psi(x) = \infty$ and*

$$\psi^3(x) \le (1-\delta)(1-\lambda)\frac{x}{\log x}\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(\lambda x/\psi(x))}\right), \tag{1}$$

*for some positive constants $\delta, \epsilon, \lambda$ and for all large $x \in \mathbb{N}$. Let the random set $A$ be defined by letting each $x \in A$, independently of the other integers and with probability $p_x = 1/\phi(x)$.*

*Then, almost surely, there is no complement $B$ of $A$ whose counting function $B(x)$ satisfies $B(x) \le \psi(x)$, infinitely often*

The result on the "prime-like" sets is the following.

**Corollary 1** *Define the random set $A \subseteq \{4, 5, 6, \ldots\}$ by* $\mathbf{Pr}[x \in A] = 1/\log x$, *independently for all $x \ge 4$. Then, almost surely, there is no complement $B$ of $A$ with counting function satisfying*

$$\liminf_{x\to\infty} \frac{B(x)}{\log^2 x} < 1. \tag{2}$$

Note that such a random set $A$ has $A(x) \sim x/\log x$, almost surely.
**Proof of Corollary 1** Let $\phi(x) = \log x$ in Theorem 2. Take $\delta = \lambda = 1/2$, for example. Then condition (1) becomes

$$\psi^3(x) \le \frac{1}{4}\frac{x}{\log x}\exp\left(\frac{-(1+\epsilon)\psi(x)}{\log(\frac{1}{2}x/\psi(x))}\right). \tag{3}$$

Take $\psi(x) = C_0 \log^2 x$, for a positive constant

$$C_0 < (1+\epsilon)^{-1}. \tag{4}$$

Then the right hand side of (3) is $\gg x^\alpha$, for some positive constant $\alpha$. Thus condition (3) is satisfied for all large $x$. By Theorem 2 then, almost surely, there is no $B$, which satisfies $B(x) \le \psi(x)$ infinitely often, and which complements $A$. Since $\epsilon > 0$ is arbitrary in (4) the Corollary is proved. $\square$

Theorem 2 can give analogous results for the case of $\phi(x) \ll \log^k x$, for any constant $k$. But it provides no useful information if $\phi(x)$ grows like a power of $x$.

# 2 Proofs

## 2.1 Bounds for large deviations

In the proof of Theorem 1 we shall make use of the following bound on Large Deviations of random variables (r.v.'s), which are sums of independent r.v.'s.

**Proposition 1** (Chernoff [1, p. 239]) *If $Y = X_1 + \cdots + X_k$, and the $X_j$ are independent indicator random variables (that is $X_j \in \{0,1\}$), then for all $\epsilon > 0$*

$$\mathbf{Pr}\left[|Y - \mathbf{E}Y| > \epsilon \mathbf{E}Y\right] \le 2e^{-c_\epsilon \mathbf{E}Y},$$

*where $c_\epsilon > 0$ is a function of $\epsilon$ alone*

$$c_\epsilon = \min\{-\log\left(e^\epsilon (1+\epsilon)^{-(1+\epsilon)}\right), \epsilon^2/2\}.$$

One uses the Chernoff bound in the following manner. The r.v. $Y$ of Proposition 1 is a quantity which, we have already proved, is on the average in a certain desired range. An invocation of the Proposition then gives us an estimate on the probability that $Y$ deviates from this range.

Assume now that, as is usually the case, we have to do this for more than one variable, say for the variables $Y_1, \ldots, Y_n$, all of them sums of independent indicator r.v.'s. Typically we only need Proposition 1 with $\epsilon = Const.$, say $\epsilon = 1/10$. We then have

$$\mathbf{Pr}\left[\exists j \ : \ |Y_j - \mathbf{E}Y_j| > \epsilon \mathbf{E}Y_j\right] \le 2\sum_{j=1}^{n} \left(e^{\mathbf{E}Y_j}\right)^{-c_\epsilon},$$

and we need the right hand side to be less than 1, which can be achieved with $\mathbf{E}Y_j \ge C \log j$ and with the constant $C$ at our disposal to make as large as we please.

We shall also use Markov's inequality,

$$\mathbf{Pr}\left[X > \lambda \mathbf{E}X\right] < \frac{1}{\lambda}, \tag{5}$$

valid for every nonnegative r.v. $X$ and for all $\lambda > 0$.

## 2.2 Proof of Theorem 1

The random set $A$ is defined by letting $x \in A$ with probability

$$\alpha_x = K\frac{\log\log x}{x}, \tag{6}$$

independently for all $x \in \mathbb{N}$, where $K > 0$ is a constant that will be determined later. (If the right hand side of (6) is not in $(0,1)$ we define $\alpha_x = 0$.)

Write $\chi_j = \mathbf{1}(j \in A)$ and notice that

$$\mathbf{E}A(x) = \mathbf{E}\sum_{j=1}^{x} \chi_j = \sum_{j=1}^{x} \alpha_j.$$

One checks easily that

$$\mathbf{E}A(x) \sim K \log x \log\log x,$$

and that with high probability the r.v.'s $A(x)$ are asymptotic to their expected values, and thus asymptotic to $K \log x \log\log x$ (since $\mathbf{E}A(x)$ is super-logarithmic, Proposition 1 applies without problems–we omit the details).

The representation function is given by

$$r(x) = \#\{p \in \mathbb{P} : x - p \in A\}.$$

Our first task is to estimate its expected value. For this we quote [2] and [6, p. 154], where it is proved that

$$\mathbf{E}r(x) \gg K \log\log x. \tag{7}$$

It is worth mentioning that the proof of (7) requires information about the distribution of primes that is finer than what the Prime Number Theorem provides. In particular, one uses the fact that, for some $\delta \in (0,1)$, the number of primes in the interval $(x, x + x^\delta)$ is $\sim x^\delta / \log x$, as $x \to \infty$.

Using Proposition 1 the estimate (7) implies that the probability of the "bad event"

$$A_x = \left\{ |r(x) - \mathbf{E}r(x)| > \frac{1}{2}\mathbf{E}r(x) \right\}$$

is bounded above

$$\mathbf{Pr}[A_x] \ll (\log x)^{-\alpha}, \tag{8}$$

for $\alpha = CK$, $C$ an absolute positive constant.

The r.v. that counts the exceptions up to $n$ (i.e. the number of integers not in $A + \mathbb{P}$) is

$$X_n = \sum_{k=1}^{n} \mathbf{1}(A_k),$$

and using (8) we obtain

$$\mathbf{E}X_n \ll \frac{n}{\log^\alpha n}.$$

Markov's inequality (5) allows us to bound the probability of the bad events

$$B_n = \{X_{2^n} > n^{-M} 2^n\},$$

where $M > 0$ is arbitrary but fixed. Applying (5) with $\lambda \sim n^{\alpha - M}$ one obtains

$$\mathbf{Pr}[B_n] \ll \frac{1}{n^{\alpha - M}}.$$

The constant $K$ is now chosen so as to yield $\alpha > M + 1$. Consequently,

$$\sum_{n\geq 1} \mathbf{Pr}[B_n] < \infty,$$

which concludes the proof, since $\sum_{n\geq n_0} \mathbf{Pr}[B_n] < 1$ for some integer $n_0$, and, therefore, with positive probability, all but finitely many of the events $B_n$ fail. □

**Remark:** Since nothing beyond properties of their distribution has been used in the preceding proof about the primes, Theorem 1 holds if the sequence of primes is replaced by any other of similar growth properties.

### 2.3 Proof of Theorem 2

Fix $B \subseteq \mathbb{N}$ with $B(x) \leq \psi(x)$ for a certain $x \in \mathbb{N}$. Notice that the number of possible intersections $B \cap [1, x]$ is at most $x^{\psi(x)}$. Our first goal is to put as many disjoint translational copies of the set $-B$ in the interval $[1, x]$ as possible, such that none of these copies intersects a certain initial interval $[1, s]$, with $s = s(x)$ to be specified later.

Let $n$ be the maximum number for which there exist integers $a_1, \ldots, a_n \in [1, x]$ such that all sets $a_j - B^{\leq a_j}$ are disjoint from each other and also disjoint from the interval $[1, s]$. The lower estimate for $n$ comes from the observation that, since it is impossible to add an extra $a_{n+1}$ to the collection $a_1, \ldots, a_n$, it must be the case that, for any $y \in [1, x]$, the set $y - B^{\leq y}$ intersects either one of $a_j - B^{\leq a_j}$, $j = 1, \ldots, n$, or the interval $[1, s]$. Let

$$G = [1, s] \cup \bigcup_{j=1}^{n} \left(a_j - B^{\leq a_j}\right)$$

and observe that $\#G \leq s + nB(x)$. Our assumption about $n$ implies that $[1, x] \subseteq G + B$. Therefore

$$x \leq (s + nB(x))B(x) \leq (s + n\psi(x))\psi(x),$$

which implies

$$n \geq \frac{x}{\psi^2(x)} - \frac{s}{\psi(x)}.$$

Now let $\lambda \in (0, 1)$ be arbitrary and define $s = \lambda x/\psi(x)$ to get the lower bound

$$n \geq (1 - \lambda)\frac{x}{\psi^2(x)}. \tag{9}$$

Assume now that $B(t) \leq \psi(t)$ for infinitely many $t \in \mathbb{N}$. Let $x$ be large and such that $B(x) \leq \psi(x)$. Define $a_1, \ldots, a_n$ as above. Let $A$ be a random set as defined in the statement of the Theorem. We estimate from above the probability that a certain $a_j$ belongs to $A + B$. We have

$$\begin{aligned}
\mathbf{Pr}[a_j \in A + B] &= 1 - \mathbf{Pr}[a_j \notin A + B] \\
&= 1 - \prod_{k \,\in\, a_j - B^{\leq a_j}} \left(1 - \frac{1}{\phi(k)}\right) \\
&\leq 1 - \left(1 - \frac{1}{\phi(s)}\right)^{\psi(x)}, \quad \text{since } [1, s] \cap (a_j - B^{\leq a_j}) = \emptyset, \\
&\leq 1 - \exp\left(-(1 + \epsilon)\frac{\psi(x)}{\phi(s)}\right),
\end{aligned}$$

using the inequality $\log(1-1/\phi(s)) \ge -(1+\epsilon)/\phi(s)$, valid for any $\epsilon>0$, since $\phi(s)$ becomes large with $x$.

The probability that *all* $a_1,\ldots,a_n$ are in $A+B$ can now be estimated easily since the events $\{a_j \in A+B\}$, $j=1,\ldots,n$, are independent, the sets $a_j - B^{\le a_j}$ being disjoint. We have

$$\begin{aligned}\mathbf{Pr}\left[a_1,\ldots,a_n \in A+B\right] &= \prod_{j=1}^{n} \mathbf{Pr}\left[a_j \in A+B\right] \\ &\le \left(1-\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(s)}\right)\right)^n.\end{aligned}$$

Define the events

$$E = \{A \text{ has a complement } B \text{ such that } B(x) \le \psi(x) \text{ infinitely often}\}$$

and, for $M, x = 1,2,\ldots$,

$$E_x^M = \{\text{there is } B \subseteq \mathbb{N} \text{ with } B(x) \le \psi(x) \text{ such that } A+B \supseteq [M,x]\}.$$

We have proved that for each $M \in \mathbb{N}$, each sufficiently large $x \in \mathbb{N}$ (to be precise, for each $x$ with $s(x) > M$) and for any fixed $B \subseteq \mathbb{N}$ with $B(x) \le \psi(x)$ the probability that $A+B \supseteq [M,x]$ is at most

$$\left(1-\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(x)}\right)\right)^n.$$

Notice also that

$$E \subseteq \bigcup_{M\ge1} \bigcap_{x\ge1} \bigcup_{y\ge x} E_y^M.$$

In other words, if $A$ has a complement $B$ which satisfies $B(x) \le \psi(x)$ infinitely often, then there is an $M \in \mathbb{N}$ and infinitely many $x \in \mathbb{N}$ such that $A+B \supseteq [M,x]$ and $B(x) \le \psi(x)$. Therefore, if for each $M \in \mathbb{N}$ we can prove $\sum_{x\ge1} \mathbf{Pr}\left[E_x^M\right] < \infty$, the desired conclusion $\mathbf{Pr}[E] = 0$ will follow, as this implies $\mathbf{Pr}\left[\bigcap_{x\ge1}\bigcup_{y\ge x} E_y^M\right] = 0$.

Let us now fix $M \in \mathbb{N}$. If we assume that for a certain $x \in \mathbb{N}$ we have $B(x) \le \psi(x)$ then the number of possibilities for the set $B \cap [1,x]$ is at most $x^{\psi(x)}$ and, using the previous estimate, we obtain for all large $x$

$$\mathbf{Pr}\left[E_x^M\right] \le x^{\psi(x)}\left(1-\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(s)}\right)\right)^n. \tag{10}$$

Taking logarithms in (10) and using the lower estimate (9) and the inequality $\log(1-y) \le -y$, for $y \ge 0$, we obtain

$$\log \mathbf{Pr}\left[E_x^M\right] \le \psi(x)\log x - (1-\lambda)\frac{x}{\psi^2(x)}\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(s)}\right).$$

Since $\psi(x) \to \infty$ it therefore suffices to have

$$\psi(x)\log x \le (1-\delta)(1-\lambda)\frac{x}{\psi^2(x)}\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(s)}\right), \tag{11}$$

for some constant $\delta > 0$, and for all sufficiently large $x$. In that case we have $\mathbf{Pr}\left[E_x^M\right] \le e^{-C\log x}$, for all large $x$ and for arbitrarily large $C>0$, which implies the finiteness of the series $\sum_{x\ge1}\mathbf{Pr}\left[E_x^M\right]$.

Rearranging (11), the probability of complementation is 0 if there are positive constants $\delta, \epsilon, \lambda$ such that for all large $x$ we have

$$\psi^3(x) \le (1-\delta)(1-\lambda)\frac{x}{\log x}\exp\left(-(1+\epsilon)\frac{\psi(x)}{\phi(\lambda x/\psi(x))}\right), \tag{12}$$

and the theorem is proved. $\square$

**Acknowledgement:** I am indebted to Prof. R.P. Kaufman for pointing out an error in a previous version of the paper.

## References

[1] N. Alon and J. H. Spencer, *The probabilistic method*, Wiley-Interscience Series in Discrete Math. and Optimization, 1992.

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

[3] P. Erdős, *On a problem of Sidon in additive number theory*, Acta Sci. Math. (Szeged) **15** (1953-54), 255-259.

[4] P. Erdős, *Problems and results in additive number theory*, Colloque sur la Théorie des Nombres (CBRM, Bruxelles), 127-137.

[5] P. Erdős and A. Rényi, *Additive properties of random sequences of positive integers*, Acta Arith. **6** (1960), 83-110.

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

[7] D. Wolke, *On a problem of Erdős in additive number theory*, preprint, (1995).