ACTA ARITHMETICA
162.3 (2014)

The additive complements of primes and Goldbach's problem

by

LI-XIA DAI and HAO PAN (Nanjing)

1. Introduction. First, we introduce the notion of additive complement. For a set $A \subseteq \mathbb{N} = \{0, 1, 2, \ldots\}$, we say that a set $B \subseteq \mathbb{N}$ is an *additive complement* of A if for every sufficiently large $n \in \mathbb{N}$, there exist $a \in A$ and $b \in B$ such that $n = a + b$, i.e., the sumset

$$A + B = \{a + b : a \in A,\, b \in B\}$$

contains all sufficiently large integers.

Morever, for $A, B \subseteq \mathbb{N}$, if $A + B$ has lower density 1, i.e., almost all positive integers n can be represented as $n = a + b$ with $a \in A$ and $b \in B$, then we say B is an *almost additive complement* of A.

The additive properties of primes are among the most fascinating topics in number theory. Let $\mathcal{P}$ denote the set of all primes. It is natural to ask what the additive complements of $\mathcal{P}$ are. By the prime number theorem, we know

$$\mathcal{P}(x) = \frac{x}{\log x}(1 + o(1)),$$

where $A(x) = |A \cap [1, x]|$ for a set $A \subseteq \mathbb{N}$. So if A is an additive complement of $\mathcal{P}$, we must have $A(x) \gg \log x$. Unfortunately, no one knows whether there exists an additive complement A of $\mathcal{P}$ satisfying $A(x) \ll \log x$. However, using probabilistic methods, Erdős [2] showed that such an additive complement exists if $\log x$ is replaced by $(\log x)^2$. That is,

- *there exists a set $A \subseteq \mathbb{N}$ with $A(x) = O((\log x)^2)$ such that the sumset $A + \mathcal{P}$ contains every sufficiently large integer.*

The main tool of Erdős' proof is the polynomial method. For more applications of the polynomial method, the readers may refer to [1].

2010 *Mathematics Subject Classification*: Primary 11P32; Secondary 05D40, 11B13, 11N36.
Key words and phrases: prime, additive complement, Janson's inequality.

DOI: 10.4064/aa162-3-1

In 1998, Ruzsa [7] improved the results of Wolke [10] and Kolountzakis [6], and showed that for every function $w(x)$ with $\lim_{x\to\infty} w(x) = \infty$, there exists an almost additive complement A of $\mathcal{P}$ with $A(x)=O(w(x)\log x)$, i.e.,

- *there exists a set $A \subseteq \mathbb{N}$ with $A(x) = O(w(x)\log x)$ such that almost all positive integers can be represented as sums of an element of A and a prime.*

In 2001, Vu [9] proved that $\mathcal{P}$ has an additive complement A of order 2 with $A(x) = O(\log x)$, i.e.,

- *there exists a set $A \subseteq \mathbb{N}$ with $A(x) = O(\log x)$ such that every sufficiently large integer n can be represented as $n = a_1 + a_2 + p$, where $a_1, a_2 \in A$ and p is a prime.*

Clearly Vu's result implies Erdős', since $(A + A)(x) \ll (A(x))^2$.

Next, let us turn to the Goldbach problem. As early as 1937, using the circle method, Vinogradov [8] solved the ternary Goldbach problem and showed that

- *every sufficiently large odd integer can be represented as the sum of three primes.*

Subsequently, with a similar discussion, Estermann [3] solved the binary Goldbach problem for almost all positive even integers:

- *almost all positive even integers can be represented as sums of two primes.*

In this note, we shall combine the results of Ruzsa and Vu with Goldbach's problem.

Theorem 1.1. *There exists a set $\mathcal{A} \subseteq \mathcal{P}$ with $\mathcal{A}(x) = O(\log x)$ such that every sufficiently large odd integer can be represented as $a_1 + a_2 + p$, where $a_1, a_2 \in \mathcal{A}$ and $p \in \mathcal{P}$.*

Theorem 1.2. *For every function $w(x)$ with $\lim_{x\to\infty} w(x) = \infty$, there exists a set $\mathcal{B} \subseteq \mathcal{P}$ with $\mathcal{B}(x) = O(w(x)\log x)$ such that almost all even positive integers can be represented as $b + p$, where $b \in \mathcal{B}$ and $p \in \mathcal{P}$.*

Note that Theorems 1.1 and 1.2 contain Vu's and Ruzsa's results, respectively. In fact, if we set $A = \mathcal{A} \cup (\mathcal{A} + \{1\})$ and $B = \mathcal{B} \cup (\mathcal{B} + \{1\})$, then clearly the sumset $A + A + \mathcal{P}$ contains all sufficiently large integers, and $B + \mathcal{P}$ has lower density 1.

The proofs of Theorems 1.1 and 1.2 will be given in the next sections.

2. Proof of Theorem 1.1. The key to our proofs is the following lemma.

LEMMA 2.1. *There exists a positive constant c_0 such that if x is sufficiently large, and if $x^{1-c_0} \le M \le x$ and $0 \le y \le x - M$, then for all even integers n with $x \le n \le x + M$, except for $O(M(\log x)^{-2})$ exceptional values, we have*

$$\sum_{\substack{n=p_1+p_2 \\ y \le p_1 \le y+M \\ x-y-M \le p_2 \le x-y+M}} 1 \ge C(n) \frac{C^* M}{(\log x)^2} \tag{2.1}$$

where

$$C(n) = \prod_{p \nmid n} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p \mid n} \left(1 + \frac{1}{p-1}\right) \tag{2.2}$$

and C^* is a constant only depending on c_0.

Proof. This lemma can be proved by the method of Jia [5], although he only discussed the case $y = x/2$. In fact, Jia proved that Lemma 2.1 holds whenever $c_0 < 5/12$. ∎

Now suppose that n is a sufficiently large odd integer. For each $x \in \mathcal{P}$, we choose x to be in $\mathcal{A}$ with probability

$$\varrho_x = \frac{\mathbf{c} \log x}{x},$$

where $\mathbf{c} > 0$ is a constant to be chosen later.

Let t_x be the binary random variable representing the choice of x, i.e., $t_x = 1$ with probability ϱ_x and 0 with probability $1 - \varrho_x$. Then with the help of the Borel–Cantelli lemma, we almost surely have

$$\mathcal{A}(n) = \sum_{x \in \mathcal{P} \cap [1,n]} t_x = O(\log n)$$

for every n, since

$$\sum_{\substack{p \le n \\ p \text{ is prime}}} \frac{\log p}{p} = \log n \, (1 + o(1)).$$

Now consider

$$Y_n = \sum_{\substack{p<n \\ p \text{ is prime}}} \sum_{\substack{i+j=n-p \\ i,j \text{ are prime}}} t_i t_j.$$

We need to prove that there exists $n_0 > 0$ such that

$$\mathbb{P}(Y_n > 0 \text{ for every } n \ge n_0) \ge 1/2.$$

Choose $0 < \epsilon < c_0/2$ and let $M = n^{1-2\epsilon}$; here and later, c_0 is a number admissible in Lemma 2.1. In the remainder of this section, the implied constants of $O(\cdot)$, $\ll$ and $\gg$ will only depend on ϵ.

Let

$$Y_n^* = \sum_{\substack{p \le n - n^{1-\epsilon} \\ p \,\text{is prime}}} \sum_{\substack{i+j=n-p \\ i,j \ge M \\ i,j \text{ are prime}}} t_i t_j.$$

Clearly $\mathbb{P}(Y_n > 0)$ is not less than $\mathbb{P}(Y_n^* > 0)$ for every n.

Next, we need a probabilistic result of Janson [4]. Let $\{J_i\}_{i\in R}$ be a set of independent random indicator variables, and let X be a collection of subsets of R. For $\alpha \in X$, define

$$I_\alpha = \prod_{i \in \alpha} J_i.$$

LEMMA 2.2 ([4, (1.9)]). *Let*

$$Y = \sum_{\alpha \in X} I_\alpha \quad \textit{and} \quad \Delta = \sum_{\substack{\alpha,\beta \in X \\ \alpha \sim \beta}} \mathbb{E}(I_\alpha I_\beta)$$

where $\alpha \sim \beta$ means $\alpha \neq \beta$ and $\alpha \cap \beta \neq \emptyset$. Then for any $\varepsilon > 0$,

$$\mathbb{P}\big(Y \le (1-\varepsilon)\mathbb{E}(Y)\big) \le \exp\bigg(-\frac{(\varepsilon \mathbb{E}(Y))^2}{2(\mathbb{E}(Y) + \Delta)}\bigg).$$

Let

$$X = \{(i,j) : i,j \ge M,\ i+j \ge n^{1-\epsilon} \text{ and } i,j,n-i-j \text{ are prime}\}.$$

Then

$$Y_n^* = \sum_{(i,j)\in X} t_i t_j = \sum_{\alpha \in X} I_\alpha,$$

where $I_\alpha = t_i t_j$ for $\alpha = (i,j)$. In view of Lemma 2.2 we only need to estimate

$$\mathbb{E}(Y_n^*) = \mathbb{E}\bigg(\sum_{\alpha \in X} I_\alpha\bigg) \quad \text{and} \quad \Delta = \mathbb{E}\Big(\sum_{\substack{\alpha,\beta \in X \\ \alpha \sim \beta}} I_\alpha I_\beta\Big).$$

By Lemma 2.1, we have

$$\begin{aligned}
\mathbb{E}\bigg(\sum_{\alpha \in X} I_\alpha\bigg) &= \mathbb{E}\bigg(\sum_{\substack{p \le n - n^{1-\epsilon} \\ p \,\text{is prime}}} \sum_{\substack{i+j=n-p \\ i,j \ge M \\ i,j \text{ are prime}}} t_i t_j\bigg) \\
&\ge \mathbb{E}\bigg(\sum_{\substack{0 \le t \le n^{1-\epsilon}/M \\ tM \le n - n^{1-\epsilon} - p < (t+1)M \\ p \,\text{is prime}}} \sum_{\substack{1 \le s \le (n-p)/M - 2 \\ i+j=n-p \\ sM \le i < (s+1)M \\ i,j \text{ are prime}}} t_i t_j\bigg)
\end{aligned}$$

$$\geq \mathbb{E}\Bigg(\sum_{\substack{0\leq t\leq n^{1-\epsilon}/M \\ 1\leq s\leq n^{1-\epsilon}/M+t-2}} \sum_{\substack{tM\leq n-n^{1-\epsilon}-p<(t+1)M \\ p \text{ is prime}}} \sum_{\substack{i+j=n-p \\ sM\leq i<(s+1)M \\ i,j \text{ are prime}}} t_i t_j \Bigg)$$

$$\gg \sum_{\substack{0\leq t\leq n^{1-\epsilon}/M \\ 1\leq s\leq n^{1-\epsilon}/M+t-2}} \frac{M}{\log M} \frac{M}{(\log M)^2} \frac{\mathbf{c}^2(\log M)^2}{sM(n^{1-\epsilon}+(t+1)M-sM)}.$$

Clearly,

$$\sum_{\substack{0\leq t\leq n^{1-\epsilon}/M \\ 1\leq s\leq n^{1-\epsilon}/M+t-2}} \frac{1}{sM(n^{1-\epsilon}+(t+1)M-sM)}$$

$$\gg \int_1^{n^{1-\epsilon}/M} \Bigg(\int_1^{n^{1-\epsilon}/M+t-2} \frac{1}{sM(n^{1-\epsilon}+tM-sM)}\, ds\Bigg) dt$$

$$\gg \int_1^{n^{1-\epsilon}/M} \frac{1}{n^{1-\epsilon}+tM} \Bigg(\int_1^{n^{1-\epsilon}/M+t-2} \Bigg(\frac{1}{sM}+\frac{1}{n^{1-\epsilon}+tM-sM}\Bigg) ds\Bigg) dt$$

$$\gg \int_1^{n^{1-\epsilon}/M} \frac{\log(n^{1-\epsilon}+tM)}{M(n^{1-\epsilon}+tM)}\, dt \gg \frac{(\log(2n^{1-\epsilon}))^2}{M^2} \gg \frac{(\log n)^2}{M^2}.$$

Thus

$$\mathbb{E}(Y_n^*) \gg \mathbf{c}^2 \log n.$$

Now we turn to Δ. One has

$$\mathbb{E}\Bigg(\sum_{\substack{\alpha,\beta\in X \\ \alpha\sim\beta}} I_\alpha I_\beta\Bigg) = \mathbb{E}\Bigg(\sum_{\substack{i\geq M \\ i \text{ is prime}}} \sum_{\substack{p_1+j_1=p_2+j_2=n-i \\ p_1,p_2\leq n-n^{1-\epsilon} \\ j_1\neq j_2,\, j_1,j_2\geq M \\ p_1,p_2,j_1,j_2 \text{ are prime}}} (t_i t_{j_1})(t_i t_{j_2}) \Bigg)$$

$$= \mathbb{E}\Bigg(\sum_{\substack{i\geq M \\ i \text{ is prime}}} t_i \sum_{\substack{p_1+j_1=p_2+j_2=n-i \\ p_1,p_2\leq n-n^{1-\epsilon} \\ j_1\neq j_2,\, j_1,j_2\geq M \\ p_1,p_2,j_1,j_2 \text{ are prime}}} t_{j_1} t_{j_2} \Bigg)$$

$$= \mathbf{c}^3 \sum_{\substack{i\geq M \\ i \text{ is prime}}} \frac{\log i}{i} \sum_{\substack{p_1+j_1=p_2+j_2=n-i \\ p_1,p_2\leq n-n^{1-\epsilon} \\ j_1\neq j_2,\, j_1,j_2\geq M \\ p_1,p_2,j_1,j_2 \text{ are prime}}} \frac{\log j_1 \log j_2}{j_1 j_2}.$$

Clearly,

$$\sum_{\substack{p_1+j_1=p_2+j_2=n-i\\ p_1,p_2\le n-n^{1-\epsilon}\\ j_1\ne j_2,\, j_1,j_2\ge M\\ p_1,p_2,j_1,j_2 \text{ are prime}}} \frac{\log j_1 \log j_2}{j_1 j_2} \le \sum_{\substack{p_1+j_1=p_2+j_2=n-i\\ p_1,p_2\le n-n^{1-\epsilon}\\ j_1,j_2\ge M\\ p_1,p_2,j_1,j_2 \text{ are prime}}} \frac{\log j_1 \log j_2}{j_1 j_2}$$
$$\le \Bigg(\sum_{\substack{p+j=n-i\\ j\ge M\\ p,j \text{ are prime}}} \frac{\log n}{j} \Bigg)^2.$$

For a set U of positive integers, using partial summation, we have

$$\sum_{\substack{M\le j\le x\\ j\in U}} \frac{1}{j} \ll \frac{U(x)}{x} - \frac{U(M)}{M} + \sum_{M<y\le x} \frac{U(y)}{y^2}.$$

Since $t/(\log t)^2$ is increasing for $t > M + e$, by the sieve method, we have

$$|\{M \le j \le y : \text{both } j \text{ and } m-j \text{ are primes}\}| \ll \frac{C(m)(y-M)}{(\log(y-M))^2} \ll \frac{C(m)y}{(\log y)^2},$$

where $C(m)$ is as in Lemma 2.1. Therefore

$$\sum_{\substack{p+j=n-i\\ j\ge M\\ p,j \text{ are prime}}} \frac{1}{j} \ll \frac{1}{n-i}\frac{C(n-i)(n-i)}{(\log(n-i))^2} + \frac{1}{M} + \sum_{M+e<y\le n-i} \frac{C(n-i)}{y(\log y)^2}$$
$$\ll \frac{C(n-i)}{\log n}.$$

It follows that

$$\mathbb{E}\Big(\sum_{\alpha\sim\beta} I_\alpha I_\beta\Big) \ll \mathbf{c}^3 \log n \sum_{\substack{M\le i\le n-n^{1-\epsilon}\\ i \text{ is prime}}} \frac{C(n-i)^2}{i}.$$

Note that

$$C(n-i)^2 \ll \prod_{p|n-i}\left(1+\frac{1}{p-1}\right)^2 \ll \prod_{p|n-i}\left(1+\frac{2}{p}\right) = \sum_{\substack{d|n-i\\ d \text{ is square-free}}} \frac{2^{\omega(d)}}{d},$$

where $\omega(d)$ denotes the number of distinct prime factors of d. Hence,

$$\sum_{\substack{M\le i\le n-n^{1-\epsilon}\\ i \text{ is prime}}} \frac{C(n-i)^2}{i} \ll \sum_{\substack{d<n-n^{1-\epsilon}-M\\ d \text{ is square-free}}} \frac{2^{\omega(d)}}{d} \sum_{\substack{M\le i\le n-n^{1-\epsilon}\\ i \text{ is prime}\\ i\equiv n \,(\mathrm{mod}\, d)}} \frac{1}{i}.$$

By the Brun–Titchmarsh theorem, we know that

$$|\{M \le i \le y : i \text{ is prime and } i \equiv n \pmod d\}| \ll \frac{y-M}{\phi(d)\log((y-M)/d)},$$

provided that $y - M \ge \frac{11}{10}d$. By partial summation, whenever

$$d \le \frac{10}{11}(n - n^{1-\epsilon} - M),$$

one finds that

$$\sum_{\substack{M \le i \le n-n^{1-\epsilon} \\ i \text{ is prime} \\ i \equiv n \,(\mathrm{mod}\, d)}} \frac{1}{i} \ll \frac{1}{n-n^{1-\epsilon}} \frac{n}{\phi(d)\log(n/d)} + \frac{1}{M} + \sum_{M<y<M+\frac{11}{10}d} \frac{2}{y^2}$$

$$+ \sum_{M+\frac{11}{10}d \le y \le n-n^{1-\epsilon}} \frac{1}{\phi(d) y \log(y/d)}$$

$$\ll \frac{1}{\phi(d)} + \frac{1}{M} + \frac{\log\log(n/d) - \log\log(M/d)}{\phi(d)}.$$

If $d \le \sqrt{M}$, then

$$\log\log(n/d) - \log\log(M/d) \le \log\log n - \log\log\sqrt{M} \ll 1.$$

If $\sqrt{M} < d \le \frac{10}{11}(n - n^{1-\epsilon} - M)$, then

$$\frac{\log\log(n/d) - \log\log(M/d)}{\phi(d)} \ll \frac{\log\log n}{d^{1-\epsilon}} \ll \frac{1}{M^{1/2-\epsilon}}.$$

Thus we have

$$\sum_{\substack{d \le \frac{10}{11}(n-n^{1-\epsilon}-M) \\ d \text{ is square-free}}} \frac{2^{\omega(d)}}{d} \sum_{\substack{M \le i \le n-n^{1-\epsilon} \\ i \text{ is prime} \\ i \equiv n \,(\mathrm{mod}\, d)}} \frac{1}{i} \ll \sum_{d<n} \frac{2^{\omega(d)}}{d}\left(\frac{1}{M^{1/2-\epsilon}} + \frac{1}{\phi(d)}\right) \tag{2.3}$$

$$\ll \frac{1}{M^{1/2-\epsilon}} \sum_{d<n} \frac{1}{d^{1-\epsilon}} + \sum_{d<n} \frac{1}{d^{2-\epsilon}} \ll \frac{n^\epsilon}{M^{1/2-\epsilon}} + O(1) = O(1).$$

On the other hand,

$$\sum_{\substack{\frac{10}{11}(n-n^{1-\epsilon}-M) < d \le n-n^{1-\epsilon}-M \\ d \text{ is square-free}}} \frac{2^{\omega(d)}}{d} \sum_{\substack{M \le i \le n-n^{1-\epsilon} \\ i \text{ is prime} \\ i \equiv n \,(\mathrm{mod}\, d)}} \frac{1}{i}$$

$$\ll \sum_{\substack{\frac{10}{11}(n-n^{1-\epsilon}-M) < d \le n-n^{1-\epsilon}-M \\ d \text{ is square-free}}} \frac{2^{\omega(d)}}{n} \sum_{\substack{M \le i \le n-n^{1-\epsilon} \\ i \text{ is prime} \\ i = n-d}} 1 \le \frac{1}{n} \sum_{\substack{i<n \\ i \text{ is prime}}} \tau(n-i),$$

where τ is the divisor function. Recall that $\tau(d) = 2^{\omega(d)}$ for d square-free. Now

$$\sum_{\substack{i<n \\ i \text{ is prime}}} \tau(n-i) = \sum_{\substack{i<n \\ i \,\text{is prime}}} \sum_{k|n-i} 1 \le 2 \sum_{\substack{i<n \\ i \,\text{is prime}}} \sum_{\substack{k\le\sqrt{n-i} \\ k|n-i}} 1$$
$$\le \sum_{k\le\sqrt{n}} \sum_{\substack{i<n \\ i \,\text{is prime} \\ i\equiv n \,(\text{mod}\, k)}} 1 \ll \sum_{k\le\sqrt{n}} \frac{n}{\phi(k)\log n}.$$

It is well known that

$$\sum_{k\le x} \frac{1}{\phi(k)} \ll \log x.$$

It follows that

$$\sum_{\substack{\frac{10}{11}(n-n^{1-\epsilon}-M)<d\le n-n^{1-\epsilon}-M \\ d \,\text{is square-free}}} \frac{2^{\omega(d)}}{d} \sum_{\substack{M\le i\le n-n^{1-\epsilon} \\ i \,\text{is prime} \\ i\equiv n \,(\text{mod}\, d)}} \frac{1}{i} \ll 1. \tag{2.4}$$

Thus we get

$$\Delta \ll \mathbf{c}^3 \log n.$$

Now

$$\frac{\mathbb{E}(Y_n^*)^2}{\mathbb{E}(Y_n^*)+\Delta} = \frac{\mathbb{E}(Y_n^*)}{1+\Delta/\mathbb{E}(Y_n^*)} \gg \frac{\mathbf{c}^2\log n}{1+\frac{\mathbf{c}^3\log n}{\mathbf{c}^2\log n}} = \frac{\mathbf{c}^2}{1+\mathbf{c}}\log n.$$

Therefore we may choose $\mathbf{c}$ sufficiently large so that

$$\frac{\mathbb{E}(Y_n^*)^2}{\mathbb{E}(Y_n^*)+\Delta} \ge 100\log n.$$

Applying Lemma 2.2, we have

$$\mathbb{P}(Y_n^*=0) \le \mathbb{P}\left(Y_n^* \le \frac{1}{2}\mathbb{E}(Y_n^*)\right) \le \exp\left(-\frac{-(\frac{1}{2}\mathbb{E}(Y_n^*))^2}{2(\mathbb{E}(Y_n^*)+\Delta)}\right)$$
$$\le \exp(-2\log n) = 1/n^2.$$

Hence for a sufficiently large n_0,

$$\mathbb{P}(Y_n^*=0 \text{ for some } n\ge n_0) \le \sum_{n\ge n_0} \mathbb{P}(Y_n^*=0) \le \sum_{n\ge n_0} \frac{1}{n^2} \le \frac{1}{2}.$$

This completes the proof of Theorem 1.1.

3. Proof of Theorem 1.2. Let c_0 be the constant appearing in Lemma 2.1 and c_1 be another fixed constant with $c_0 < c_1 < 1$.

LEMMA 3.1. *For every $0 < \epsilon < 1$, there are $K = K(\epsilon)$ and $N_0 = N_0(\epsilon)$ such that for $N > N_0$ we can always find a set*

$$B \subset [N^{c_0}, 2N^{c_0}]$$

of primes such that

$$|B| \le K \log N \tag{3.1}$$

and the set $S = \mathcal{P} + B$ satisfies

$$S(x) \ge (1-\epsilon)x \quad \textit{for all } N^{c_1} \le x \le N. \tag{3.2}$$

Proof. Let

$$K = \max\left\{\frac{10}{c_0 C^*} \log \frac{16}{\epsilon^2}, 20\right\}, \qquad M = N^{c_0},$$
$$I = [M, 2M] \cap \mathcal{P}, \qquad L = |I| = (1+o(1))\frac{M}{\log M},$$

where C^* is as in (2.1). For $x \in I$, choose x to be in B with probability

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

It suffices to show that

$$\mathbb{P}((3.2) \text{ holds}) > \mathbb{P}(|B| > K \log N). \tag{3.3}$$

Let t_x be the binary random variable representing the choice of x. Evidently,

$$\mathbb{E}(e^{|B|}) = \mathbb{E}\biggl(\exp\Bigl(\sum_{x\in I} t_x\Bigr)\biggr) = \prod_{x\in I} \mathbb{E}(e^{t_x}) = ((1-\varrho)\cdot e^0 + \varrho \cdot e^1)^L$$
$$= (1+\varrho(e-1))^L \le \exp(\varrho(e-1)L) < \exp\left(\frac{K(e-1)}{19/10}\log N\right).$$

Thus from Markov's inequality we get

$$\mathbb{P}(|B| > K\log N) = \mathbb{P}(e^{|B|} > e^{K\log N}) \le \mathbb{E}(e^{|B|})/\exp(K\log N)$$
$$\le \exp\left(\frac{K(e-1)}{19/10}\log N - K\log N\right) = N^{\frac{K(e-29/10)}{19/10}} < \frac{1}{N}.$$

Below we shall show that $\mathbb{P}((3.2) \text{ holds}) \ge N^{-1}$. Let $\eta = 1 + \epsilon/2$ and let $x_j = N/\eta^j$ for $0 \le j \le (1-c_1)\log N/\log \eta + 1$. Obviously, for $x_j \le x \le x_{j-1}$, the inequality $S(x_j) > (1-\epsilon/2)x_j$ implies that $S(x) > (1-\epsilon)x$. So it suffices to show that

$$\mathbb{P}\bigl(S(x_j) > (1-\epsilon/2)x_j \text{ for all } j\bigr) > 1/N.$$

Let

$$T(x) = x - S(x).$$

Clearly, $S(x_j) > (1-\epsilon/2)x_j$ is equivalent to $T(x_j) < (\epsilon/2)x_j$. Let

$$z(n) = \sum_{\substack{n=p_1+p_2\\ p_1\in I,\, p_2\in\mathcal{P}}} 1$$

and

$$Q_j = \left\{ M \le n \le x_j : z(n) < \frac{C^* C(n) M}{3\log^2 n} \right\}.$$

By Lemma 2.1, we have

$$\begin{aligned} |Q_j| &\le \sum_{1\le k\le x_j/M} |Q_j \cap [kM,(k+1)M]| \\ &= \sum_{1\le k\le x_j/M} O\left(\frac{M}{(\log(kM))^2}\right) = O\left(\frac{x_j}{(\log M)^2}\right). \end{aligned}$$

For a given $n \in [2M, x_j] \setminus Q_j$, we can have $n \notin S$ only in those cases where none of the primes $p_1 \in I$ for which $n - p_1$ is also prime happens to be in B. The probability of this event is

$$\mathbb{P}(n \notin S) = (1-\varrho)^{z(n)} \le \exp(-\varrho z(n)).$$

Note that

$$\varrho z(n) \ge \frac{K\log N}{2L} C^* C(n) \frac{M}{3\log^2 n} \ge \frac{C^* C(n) K M}{6L\log N} \ge \frac{C^* c_0 K}{10},$$

where we set

$$C(n) \ge \prod_{p\ge 3 \text{ is prime}} \left(1 - \frac{1}{(p-1)^2}\right) > \frac{6}{10}.$$

Hence,

$$\mathbb{P}(n \notin S) \le \exp(-c_0 C^* K/10). \tag{3.4}$$

Thus the expectation of $T(x_j)$ satisfies

$$\begin{aligned} \mathbb{E}(T(x_j)) &\le \exp(-c_0 C^* K/10) x_j + 2M + O(x_j (\log M)^{-2}) \\ &\le 2\exp(-c_0 C^* K/10) x_j, \end{aligned} \tag{3.5}$$

by noting that $x_j \ge N^{c_1}$. From Markov's inequality,

$$\mathbb{P}\big(T(x_j) \ge (\epsilon/2)x_j\big) \le \frac{\mathbb{E}(T(x_j))}{(\epsilon/2)x_j} < \frac{4}{\epsilon}\exp\left(-\frac{c_0 C^* K}{10}\right) \le 1 - \frac{1}{\eta}, \tag{3.6}$$

since

$$K \ge \frac{10}{c_0 C^*} \log \frac{4}{\epsilon(1-\eta^{-1})}. \tag{3.7}$$

So letting $J = \lfloor (1-c_1)\log N/\log\eta + 1\rfloor$, we get

$$\mathbb{P}\big(S(x_j) > (1-\epsilon/2)x_j \text{ for all } j\big) \ge \prod_{j=0}^{J} \mathbb{P}(S(x_j) > (1-\epsilon/2)x_j) \ge \eta^{-J-1} \ge \frac{1}{\eta^2 N^{1-c_1}} \ge \frac{1}{N}, \tag{3.8}$$

where $\lfloor\alpha\rfloor = \max\{z \in \mathbb{Z} : z \le \alpha\}$. Thus the proof of Lemma 3.1 is complete. ∎

Lemma 3.2. *For every $\epsilon > 0$, let $K = K(\epsilon)$ and $N_0 = N_0(\epsilon)$ be as in Lemma 3.1. Then there exists a set $B \subseteq \mathcal{P}$ such that for every $x > N_0$, we have*

$$B(x) \le \frac{2K}{c_0c_1(1-c_1)}\log x$$

and the sumset $S = \mathcal{P} + B$ satisfies

$$S(x) \ge (1-\epsilon)x.$$

Proof. Let $N_{i+1} = \lfloor N_i^{1/c_1}\rfloor + 1$ for $i \ge 0$. Applying Lemma 3.1 to each $N = N_i$, we get a set $B_i \subset [N_i^{c_0}, 2N_i^{c_0}]$ with $|B_i| \le K\log N_i$ satisfying $S_i(x) \ge (1-\epsilon)x$ for any $N_i^{c_1} \le x \le N_i$, where $S_i = \mathcal{P} + B_i$. We put

$$B = \bigcup B_i.$$

Suppose that $x \in [N_i^{c_1}, N_i]$. Clearly, for $S = \mathcal{P} + B$, we have

$$S(x) \ge S_i(x) \ge (1-\epsilon)x.$$

Let

$$k = \min\{i : N_i^{c_0} > x \text{ and } N_{i-1}^{c_0} \le x\}.$$

Note that

$$\log N_{i-1} \le c_1 \log N_i,$$

i.e., $\log N_i$ grows exponentially. So we get

$$B(x) \le \sum_{0\le i\le k} |B_i| \le K(\log N_0 + \log N_1 + \cdots + \log N_k) \le K\log N_k \sum_{j=0}^{k} c_1^j \le \frac{2K\log(x^{1/c_0c_1})}{1-c_1}. \ ∎$$

Now we are ready to prove Theorem 1.2.

Proof of Theorem 1.2. Let $\epsilon_i = 1/(i+1)$. Let B_i be the set satisfying the requirements of Lemma 3.2 for $\epsilon = \epsilon_i$. Let $N_1 = N_0(\epsilon_1)$. Since $w(x) \to \infty$ as $x \to \infty$, for every $i \ge 2$, let $N_i \ge \max\{N_0(\epsilon_i), e^{2N_{i-1}/\epsilon_i}\}$ be an integer such that

$$\frac{2}{c_0 c_1(1-c_1)} \sum_{j=1}^{i+1} K(\epsilon_j) \le w(x)$$

for every $x \ge N_i$. Let

$$\mathcal{B} = \bigcup_{i=1}^{\infty} (B_i \cap [N_{i-1}, N_{i+1}]).$$

Note that for any $N_i < x \le N_{i+1}$, we have

$$\begin{aligned}
&|\{n \le x : n = p + b,\ p \in \mathcal{P},\ b \in \mathcal{B}\}| \\
&\qquad \ge |\{n \le x : n = p + b,\ p \in \mathcal{P},\ b \in B_i \cap [N_{i-1}, N_{i+1}]\}| \\
&\qquad > |\{n \le x : n = p + b,\ p \in \mathcal{P},\ b \in B_i\}| - N_{i-1} \cdot \mathcal{P}(x) \\
&\qquad \ge (1-\epsilon_i)x - N_{i-1} 2x/\log x \ge (1-2\epsilon_i)x.
\end{aligned}$$

So $\mathcal{P} + \mathcal{B}$ has lower density 1. Also, for every $N_i < x \le N_{i+1}$,

$$\mathcal{B}(x) \le \sum_{j=1}^{i+1} |B_j \cap [1,x]| \le \left(\frac{2}{c_0 c_1 (1-c_1)} \sum_{j=1}^{i+1} K(\epsilon_j) \right) \log x \le w(x) \log x. \quad \blacksquare$$

Acknowledgements. The authors are grateful to the anonymous referee for his/her useful comments and helpful suggestions. The second author also thanks his colleague, Dr. Rongli Liu, for her discussions on Vu's proof.

The authors are supported by National Natural Science Foundation of China (Grant No. 11271185).

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] T. Estermann, *On Goldbach's problem: proof that almost all even positive integers are sums of two primes*, Proc. London Math. Soc. 44 (1938), 307–314.
[4] S. Janson, *Poisson approximation for large deviations*, Random Structures Algorithms 1 (1990), 221–230.
[5] C.-H. Jia, *Three primes theorem in a short interval (VII)*, Acta Math. Sinica 10 (1994), 369–387.
[6] M. N. Kolountzakis, *On the additive complements of the primes and sets of similar growth*, Acta Arith. 77 (1996), 1–8.
[7] I. Ruzsa, *On the additive completion of primes*, Acta Arith. 86 (1998), 269–275.
[8] I. M. Vinogradov, *The representation of an odd number as a sum of three primes*, Dokl. Akad. Nauk SSSR 16 (1937), 139–142.
[9] V. H. Vu, *High order complementary bases of primes*, Integers 2 (2002), A12, 5 pp.

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

Li-Xia Dai
School of Mathematical Sciences
Nanjing Normal University
Nanjing 210046
People's Republic of China
E-mail: lilidainjnu@163.com

Hao Pan (corresponding author)
Department of Mathematics
Nanjing University
Nanjing 210093
People's Republic of China
E-mail: haopan79@yahoo.com.cn

Received on 4.5.2011
and in revised form on 15.7.2013 (6686)