Acta Math. Hungar.
106 (3) (2005), 197–226.

ON THIN SETS OF PRIMES EXPRESSIBLE AS SUMSETS

E. S. CROOT III (Atlanta) and C. ELSHOLTZ (Egham)

Abstract. Whether or not there exist sets of integers A and B, each with at least two elements, such that $A+B$ coincides with the set of primes for sufficiently large elements, remains an open problem. There has been recent progress, however, showing that the counting functions $A(x)$ and $B(x)$ must both have size $x^{1/2+o(1)}$.

We show in this paper that further progress can be expected from the structure theory of sumsets. As a first step towards this, we examine sumsets of three sets A, B, C, where each has at least two elements, and $A+B+C$ consists entirely of primes. First we show that, assuming the Hardy–Littlewood conjecture, there exist sets of integers A, B, C, each having at least two elements, with $A+B+C$ consisting entirely of primes, and $(A+B+C)(x) \gg x/\log^3 x$, and where $A+B$ contains at most 3 elements. Thus, there exist "not very thin" sets of primes that can be expressed as a sumset of three sets. The main result in the paper is a certain "inverse theorem": We show that if A, B, C each have at least two elements, $A+B+C$ consists entirely of primes with $(A+B+C) \gg x/\log^\kappa x$, and if A, B, C are what we call a "regular triple", then either $A+B$, $B+C$ or $A+C$ must have at most κ elements. We use many different methods to prove this, including sieve methods, the probabilistic method, and a variety of other combinatorial methods.

1. Introduction

We will use the following notation. Given a set of positive integers S, let $S(x)$ denote the number of elements in S that are $\leqq x$, and let $|S|$ denote the total number of elements of S. Given two sets of positive integers A and B, denote the *sumset* $\{a+b : a \in A,\, b \in B\}$ by $A+B$; and so, the number of elements in $A+B$ that are $\leqq x$ will be $(A+B)(x)$. For a finite set of integers J, and integers $q \geqq 2$ and r, let $J(r,q)$ denote the set of elements of J which are $\equiv r \pmod q$. We will also use Vinogradov's notation: The statements "$f(x) \ll g(x)$" and "$g(x) \gg f(x)$" are both equivalent to "$f(x)$

Key words and phrases: primes, sumsets.
2000 Mathematics Subject Classification: 11P70, 11N36.

$= O\big(g(x)\big)$"; and, we will use "$f(x) \ll_y g(x)$" to indicate that the implied constant in the "O" depends on a parameter y. Finally, by the statement $f(x) \sim g(x)$ we mean that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

An old conjecture of Ostmann, which is sometimes called the 'Inverse Goldbach Problem', asks whether there is an additive decomposition of the primes, with at most finitely many exceptions (see [14], p. 13); that is, do there exist sets of positive integers A and B, each with at least two elements, such that

$$\text{for } n > x_0 \text{ (for some } x_0), \quad n \in A + B \iff n \text{ is prime.}$$

Even though this question has withstood attack by several mathematicians, there has been much recent progress. For instance, Wirsing in [17], Pomerance, Sárközy and Stewart in [15], Hofmann and Wolke in [10] and Bshouty and Bshouty in [1] have shown that if such a decomposition exists, then one has the following estimates on the counting functions:

$$\frac{x}{\log x} \ll A(x)B(x) \ll x.$$

Elsholtz has shown in [4] that

$$\frac{\sqrt{x}}{\log^5 x} \ll \min\big(A(x), B(x)\big) \leqq \max\big(A(x), B(x)\big) \ll \sqrt{x}\log^4 x,$$

which improves on another result of Hofmann and Wolke from [10]. Furthermore, Elsholtz uses these results to show that B (or A) cannot be the sum of two other sets, each with at least two elements; that is, Elsholtz has solved a ternary analogue of the above conjecture of Ostmann.

It seems conceivable that sieve methods alone will not solve the Ostmann problem, but that some additional insight into the structure of sumsets is needed. This paper is a step towards this.

In his proof of the above mentioned result, Elsholtz makes strong use of the fact that

$$(A + B + C)(x) \gg \frac{x}{\log x}.$$

This leads one to wonder whether this constraint can be weakened somewhat. To be more specific: Do there exist $\kappa > 1$ and sets of positive integers A, B,

C, each with at least two elements, such that $A+B+C$ is a set of primes with

$$(A+B+C)(x) \gg \frac{x}{\log^{\kappa} x}? \tag{1}$$

The answer to this question is yes, assuming the Hardy–Littlewood conjecture, which can be used to give solutions for every $\kappa \geqq 3$. Before we show how, we give here the form of the conjecture we will need (see [9]):

Hardy–Littlewood Conjecture. *Suppose that $a_1 < a_2 < \cdots < a_k$ is a sequence of integers such that the polynomial $(x+a_1)(x+a_2)\cdots(x+a_k)$ has no fixed prime divisors. Then,*

$$\#\{n \leqq x : n+a_1, n+a_2, \ldots, n+a_k \text{ are all prime}\} \sim C(a_1,\ldots,a_k)\frac{x}{\log^k x},$$

where $C(a_1,\ldots,a_k)$ is some constant which depends only on $a_1,\ldots,a_k$.

Now, suppose that $A = B = \{1,7\}$, and let C be the set of all positive integers n such that $n+2$, $n+8$, $n+14$ are all prime. Then, $A+B+C$ consists entirely of primes, and since $(x+2)(x+8)(x+14)$ has no fixed prime divisors, assuming the Hardy–Littlewood conjecture we get that

$$C(x) \sim C(2,8,14)\frac{x}{\log^3 x};$$

and so,

$$(A+B+C)(x) \gg \frac{x}{\log^3 x},$$

which means that our question above has an affirmitive answer for all $\kappa \geqq 3$.

So, if we are to have any hope of extending Elsholtz's work to show that there are no triples A, B, C where $A+B+C$ contains *many* primes (but not almost all primes), we have to account for the above "obstruction" arising from the Hardy–Littlewood conjecture. The following are all the cases where one can apply the Hardy–Littlewood conjecture to construct sets A, B, C such that (1) holds:

$$\text{either} \quad |A+B| \leqq \kappa, \quad \text{or} \quad |A+C| \leqq \kappa, \quad \text{or} \quad |B+C| \leqq \kappa. \tag{2}$$

This now leads us to the following general conjecture:

Conjecture 1. *Suppose A, B, C are sets of positive integers with at least two elements each, such that $A+B+C$ consists entirely of primes. If*

$$(A+B+C)(x) \gg_\kappa \frac{x}{\log^\kappa x},$$

then (2) *holds.*

In this paper we do not quite prove this conjecture, although we believe that it is true. One additional, technical assumption about the sets A, B, C is needed for our proof; basically, we need that there are not "too many" primes p which have "too many" solutions $p=a+b+c$, $a \in A$, $b \in B$, $c \in C$. To state this technical assumption, we need the following definition:

Definition. For a given collection of sets $A_1, A_2, \dots, A_k$ let $r(n; A_1, \dots, A_k)$ denote the number of solutions to

$$n = a_1 + \dots + a_k, \quad a_i \in A_i \quad \text{for all} \quad i = 1, 2, \dots, k.$$

We say that the collection of sets $A_1, \dots, A_k$ is *regular* if and only if for every $\varepsilon > 0$, there exists $D > 0$ such that for x sufficiently large,

$$\sum_{n \in S} r(n; A_1, \dots, A_k) < \varepsilon A_1(x) \cdots A_k(x), \tag{3}$$

where

$$S = \left\{ n \in A_1 + \dots + A_k : n \leqq x, \ r(n; A_1, \dots, A_k) > \log^D x \right\}. \tag{4}$$

One easy consequence of the fact that $A_1, \dots, A_k$ is regular is the following:

Lemma 1. *Suppose $A_1, \dots, A_k$ is regular. Then, there exists a constant $E > 0$ such that for x sufficiently large,*

$$A_1(x) A_2(x) \cdots A_k(x) < (\log^E x)(A_1 + \dots + A_k)(x).$$

A proof of this lemma can be found in Section 4.

Our main theorem is as follows:

Theorem 1 (Main Theorem). *Conjecture* 1 *holds if we assume A, B, C is a regular triple; that is, if A, B, C is a regular triple of sets of positive integers such that $A+B+C$ is a set of primes, and*

$$|A|, |B|, |C| \geqq 2, \quad \textit{and} \quad (A+B+C)(x) \gg \frac{x}{\log^\kappa x}, \tag{5}$$

then (2) *holds.*

NOTE. The conclusion of this theorem can *possibly* be proved under a weaker notion of *regularity:* One can maybe replace the "$\log^D x$" in (4) with "$\exp\left(\log^{1-o(1)} x\right)$", and still have the theorem hold. This would require substantial modifications of many parts of the argument, including Propositions 3 and 4, and Corollary 1.

Perhaps the most interesting feature of the theorem is the many different ingredients which are used to prove it (it looks like a problem tailor-made for a single application of some sieve method), and the way they all fit together. These include: the Large Sieve, Brun's Sive, Gallagher's Sieve, the "probabilistic method" and regularity principles (which are used to prove Proposition 3), translation invariant principles (which appear in Proposition 4 and Lemma 10), and certain "maximality" or "local-global" principles (which appear in the proof of Lemma 9).

The basic idea of the proof (of Theorem 1) is as follows: We will prove the contrapositive of the theorem by first assuming that (2) fails to hold. Through a combination of Propositions 1 and 2 (which appear in the next section of the paper) and some basic combinatorial arguments, we will find subsets of $A\cap[1,x]$, $B\cap[1,x]$ and $C\cap[1,x]$, call these subsets $\hat{A}$, $\hat{B}$ and $\hat{C}$, which will have certain usable properties. At this point, the proof will break down into two cases, with Case 1 being where $\min\left(|\hat{A}|,|\hat{B}|\right)>\kappa$ and Case 2 where this min is $\leqq\kappa$. The most difficult and important case will be Case 2; and for this case, we will construct subsets of $\hat{A},\hat{B}$ and $\hat{C}$, call these subsets S,L^* and C^*, such that the following inequalities hold:

$$(6)\qquad |S|\leqq\kappa,\quad |S+L^*+C^*|\geqq\frac{|L^*|\,|C^*|}{2}>\frac{A(x)B(x)C(x)}{\log^E x},$$

for some $E>0$, and

$$\frac{\sqrt{x}}{\log^{6+\kappa}x}\ll|L^*|,|C^*|\ll\sqrt{x}\log^6 x.$$

Then, we will show that most triples $(a,b,c)\in S\times L^*\times C^*$ have the property that, for any integer $r\geqq1$ and some integer k (and x sufficiently large), the numbers

$$a+b+c+k,\ a+b+c+2k,\ \dots,\ a+b+c+rk$$

have very few prime divisors in certain "long intervals". This result will follow by showing that the sets L^* and C^* are approximately "locally translation invariant", meaning that for 'many' primes $p\leqq\sqrt{x}$, the residue classes modulo p occupied by L^* will be almost exactly the same as those occupied by $L^* + k, L^* + 2k, \ldots,$ and $L^* + rk$; the same will hold for C^*. The method used to prove this will involve combining very precise arithmetic information about the sets L^* and C^* together with a variant of Gallagher's Larger Sieve, and will be the subject of Lemma 10 within the proof of Proposition 4.

Next, using Brun's upper bound sieve, we will show that the number of integers $n \leqq x$ where $n + k, n + k, \ldots, n + rk$ all have such few prime divisors in these "long intervals" is $\ll x \log^{-r/2} x$; and so, from this and (6) we will deduce

$$(A + B + C)(x) \ll |S + L^* + C^*| \log^E x \ll x \log^{-r/2+E} x = o(x \log^{-\kappa} x),$$

for $r > 2(\kappa + E)$. This will contradict the hypothesis of Theorem 1, and so the theorem will follow.

Although we indicated earlier how the Hardy–Littlewood conjecture can be used to produce sets A, B, C which satisfy the hypotheses and conclusion of this theorem for $\kappa \geqq 3$, we can give a weaker, unconditional result. Besides the sharpness of the inequalities obtained through the Hardy–Littlewood conjecture, this result is also weaker in that it only holds for a fixed x.

THEOREM 2. *Given integers $1 < \kappa_1 < \kappa_2$, for all sufficiently large x there exist sets of positive integers $A, B, C \subseteqq \{1, 2, \ldots, x\}$, with*

$$|A| = \kappa_1, \quad |B| = \kappa_2, \quad |C| \geqq |B| \geqq |A|,$$

such that $A + B + C$ consists entirely of primes, and

$$|A + B + C| > c_{\kappa_1,\kappa_2} \frac{x}{\log^{\kappa_1\kappa_2} x},$$

where c_{κ_1,κ_2} is some constant depending only on κ_1 and κ_2.

2. Proof of the Main Theorem (Theorem 1)

We will prove the contrapositive of this theorem. So, let us suppose that (2) fails to hold; that is, $|A + B|, |A + C|, |B + C| > \kappa$.

In our proof, we will first require a result that is a slight generalization of a result of Elsholtz [4], as well as a result which allows us to extract subsets $\hat{A} \subseteqq A \cap [1, x]$, $\hat{B} \subseteqq B \cap [1, x]$, and $\hat{C} \subseteqq C \cap [1, x]$, such that $\hat{A} + \hat{B} + \hat{C} \subset (\sqrt{x}, 2x)$. These first two propositions are as follows:

PROPOSITION 1. *If $F, G \subseteqq \{1, 2, \dots, x\}$, with $1 \leqq \delta < |F| \leqq |G|$, such that $F+G$ consists entirely of primes in $\big(\sqrt{x}, 2x\big)$, and if $|F|\,|G| \gg x/\log^{\delta} x$, then*

$$\frac{\sqrt{x}}{\log^{\delta+6} x} \ll |F| \leqq |G| \ll \sqrt{x}\log^6 x.$$

NOTE. The constant 6 can certainly be improved here, but such an improvement does not much affect the quality of the main result in this paper.

PROPOSITION 2. *If A, B, C is a regular triple, then there exist subsets $\hat{A} \subseteqq A \cap [1,x]$, $\hat{B} \subseteqq B \cap [1,x]$, $\hat{C} \subseteqq C \cap [1,x]$, such that*

$$|\hat{A}| \sim A(x), \quad |\hat{B}| \sim B(x), \quad \text{and} \quad |\hat{C}| \sim C(x), \tag{7}$$

where

$$|\hat{A}+\hat{B}+\hat{C}| \sim (A+B+C)(x) \gg \frac{x}{\log^{\kappa} x},$$

and $\hat{A}+\hat{B}+\hat{C} \subset \big(\sqrt{x}, \infty\big)$.

For x sufficiently large, we may assume that

$$|\hat{A}|, |\hat{B}|, |\hat{C}| \geqq 2, \quad \text{and} \quad |\hat{A}+\hat{B}|, |\hat{A}+\hat{B}|, |\hat{B}+\hat{C}| > \kappa. \tag{8}$$

The first inequality holds since

$$|\hat{A}| \sim A(x) \geqq 2, \quad |\hat{B}| \sim B(x) \geqq 2, \quad \text{and} \quad |\hat{C}| \sim C(x) \geqq 2;$$

and the second inequaltiy holds for similar reasons.

For a given x, suppose that, without loss of generality, $A(x), B(x) \leqq C(x)$. Consider the two sets $\hat{A}+\hat{B}$ and $\hat{C}$, and let F be the set with the smaller number of elements, and G be the set with the larger number of elements. We will show that these two sets F and G satisfy the hypotheses of Proposition 1, and we will use the conclusion of this proposition to show that $|\hat{A}+\hat{B}|$ is "large", which will be important in later arguments.

We first claim that $|G| \geqq |F| > \kappa$ for x sufficiently large. To see this, we note that

$$|\hat{C}|^3 \sim C(x)^3 \geqq A(x)B(x)C(x) \sim |\hat{A}|\,|\hat{B}|\,|\hat{C}| \geqq |\hat{A}+\hat{B}+\hat{C}| \gg \frac{x}{\log^{\kappa} x}.$$

Thus, $|\hat{C}| > \kappa$ for x sufficiently large; and, $|\hat{A}+\hat{B}| > \kappa$, by (8). It follows then that $|F|, |G| > \kappa$ for x sufficiently large.

We also have that $F+G \subseteqq \left(\sqrt{x},\infty\right)$, since $\hat{A}$, $\hat{B}$, $\hat{C}$ satisfy the conclusion to Proposition 2. Thus, F and G satisfy the hypotheses, and therefore the conclusion of Proposition 1 with $\delta=\kappa$. Thus,

$$\frac{\sqrt{x}}{\log^{6+\kappa} x} \ll |\hat{A}+\hat{B}|, |\hat{C}| \ll \sqrt{x}\log^6 x. \tag{9}$$

Between the sets $\hat{A}$ and $\hat{B}$, let S be the one with the smaller number of elements, and let L be the set with the larger number of elements. We now distinguish two cases: Case 1 is where $|S|>\kappa$, and Case 2 is where $|S| \leqq \kappa$.

To prove (the contrapositive of) the Main Theorem in Case 1, we consider the two sets $L+\hat{C}$ and S, and let F be the set with the smaller number of elements, and G be the one with the larger number of elements. (Note that the sets F and G have now changed from how we defined them before.) These sets F and G satisfy the hypotheses of Proposition 1 with $\delta=\kappa$, since

$$|S|>\kappa \quad \text{and} \quad |L+\hat{C}| \geqq |L| \geqq |S| > \kappa \ \Rightarrow\ |G| \geqq |F| > \kappa,$$

and since F and G satisfy the other hypotheses of the proposition. As in (9), we deduce from this that

$$\frac{\sqrt{x}}{\log^{6+\kappa} x} \ll |L+\hat{C}|, |S| \ll \sqrt{x}\log^6 x.$$

From this and (9) we deduce

$$A(x), B(x), C(x) \gg \frac{\sqrt{x}}{\log^{6+\kappa} x}.$$

Now, since A, B, C is a regular triple, this bound and Lemma 1 give

$$(A+B+C)(x) > x^{3/2-o(1)},$$

which is absurd.

We now consider Case 2, which is where $|S| \leqq \kappa$. For this case we will have from (9) that

$$\frac{\sqrt{x}}{\log^{6+\kappa} x} \ll |L|, |\hat{C}| \ll \sqrt{x}\log^6 x. \tag{10}$$

We need the following proposition to find subsets of L and $\hat{C}$ with usable properties.

PROPOSITION 3. *There exists a constant* $D > 0$ *such that if* x *is sufficiently large, and if* $|S| \leqq \kappa$ *(Case* 2*), then there exist subsets* $L^* \subset L$ *and* $C^* \subset \hat{C}$ *with*

$$|L^*| > \frac{|L|}{\log^D x}, \quad \textit{and} \quad |C^*| > \frac{|\hat{C}|}{\log^D x}, \tag{11}$$

such that

$$|L^*|\,|C^*| \leqq 2|L^* + C^*|. \tag{12}$$

Let $s_1, s_2 \in S$, with $s_2 > s_1$, be any two integers, set $k = s_2 - s_1$, and let

$$L^{\#} = L^* + s_1 = \{\ell + s_1 : \ell \in L^*\}.$$

Then,

$$L^{\#} + C^* = \big\{\ell + c + s_1 : (\ell, c) \in L^* \times C^*\big\}$$

consists entirely of primes, and so does $L^{\#} + C^* + k$.

We will need the following proposition and its corollary to unlock the structure of the set $L^{\#} + C^*$:

PROPOSITION 4. *Let*

$$Q = \max\big(|L^{\#}|, |C^*|\big) \gg \frac{x^{1/2}}{\log^{\kappa+D+6} x},$$

by (11) *and* (10). *Then, for any integer* $j \geqq 1$ *we will have*

$$\sum_{p \leqq Q} (\log p)\#\big\{(\ell, c) \in L^{\#} \times C^* : \ell + c + jk \equiv 0 \pmod{p}\big\}$$

$$= O\big(j|L^{\#}|\,|C^*| \log\log x\big).$$

COROLLARY 1. *There exists a constant* $c > 0$ *such that all but at most* $|L^{\#} + C^*|/2$ *of the elements* $n \in L^{\#} + C^*$ *have*

$$\sum_{j=1}^{r} \sum_{\substack{p \leqq Q \\ p \mid n+jk \\ p \text{ prime}}} \log p < cr^2 \log\log x. \tag{13}$$

One more lemma will establish the Main Theorem:

Lemma 2. *For x sufficiently large, there are at most $x \log^{-r/2} x$ integers $n \leqq x$ which satisfy* (13).

We have from Proposition 2, Corollary 1, Lemma 2, and Proposition 3 that for $r = 4D + 2\kappa + 2$ and x sufficiently large,

$$(A+B+C)(x) \leqq 2|S+L+\hat{C}| \leqq 2\kappa |L|\,|\hat{C}|$$

$$< 2\kappa |L^*|\,|C^*| \log^{2D} x \leqq 4\kappa |L^* + C^*| \log^{2D} x = 4\kappa |L^{\#} + C^*| \log^{2D} x$$

$$\leqq 8\kappa (\log^{2D} x) \# \big\{ n \in L^{\#} + C^* : n \text{ satisfies (13)} \big\}$$

$$\leqq 8\kappa (\log^{2D} x) \# \big\{ n \leqq 3x : n \text{ satisfies (13)} \big\}$$

$$\leqq 8\kappa (\log^{2D} x) \frac{3x}{\log^{2D+\kappa+1}(3x)} \ll_\kappa \frac{x}{\log^{\kappa+1/2} x}.$$

which contradicts (5), and so the theorem is proved.

3. Proof of Theorem 2

The proof is based on a twofold application of a counting argument due to Erdős, Stewart and Tijdeman [6]; compare also Lemma 6 in Pomerance, Sárközy and Stewart [15].

Lemma 3. *Let τ be a positive integer. Let $x > x_\tau$ be a sufficiently large positive integer and let T be a non-empty subset of $\{1, \dots, x\}$. Then there exists a set $S \subset T$ and a set of non-negative integers A such that $A + S \subset T$, and*

$$|S| \geqq \frac{\binom{|T|}{\tau}}{\binom{x-1}{\tau-1}}, \qquad |A| = \tau.$$

Since we want to prescribe the number of elements of two sets A and B we apply this lemma once again to the set S. This gives

Corollary 2. *Let κ_1, κ_2 denote positive integers. Let $x > x_{\kappa_1,\kappa_2}$ be a sufficiently large positive integer and let T be a non-empty subset of $\{1, \dots, x\}$. Let*

$$R = \frac{\binom{|T|}{\kappa_1}}{\binom{x-1}{\kappa_1-1}}.$$

Then there exists a subset $C \subset T$ and sets of non-negative integers A, B such that

$$A+B+C \subset T, \quad |C| \geqq \frac{\binom{R}{\kappa_2}}{\binom{x-1}{\kappa_2-1}}, \quad |A| = \kappa_1, \quad |B| = \kappa_2.$$

It is obvious that one could iterate this argument. We resist doing this since we concentrate on ternary problems.

Now let T denote the set of primes in $[1,x]$. Recall that by the prime number theorem with error term (see [11], §54)

$$|T| = \frac{x}{\log x} + \frac{x}{(\log x)^2} + O\left(\frac{x}{(\log x)^3}\right).$$

For large x we have that $|T| - \kappa_1 > \frac{x}{\log x}$. Hence it follows (as in the proof of Theorem 6 in [15]) that

$$R \geqq \frac{\frac{1}{\kappa_1!}\left(\frac{x}{\log x} + \frac{x}{2(\log x)^2}\right)^{\kappa_1}}{\frac{1}{(\kappa_1-1)!}x^{\kappa_1-1}} \geqq \frac{x}{\kappa_1(\log x)^{\kappa_1}} + \frac{x}{2(\log x)^{\kappa_1+1}}. \tag{14}$$

For the second application of the argument we observe that for large x we have $R - \kappa_2 > \frac{x}{\kappa_1(\log x)^{\kappa_1}}$. The argument then gives

$$|C| \geqq \frac{\frac{1}{\kappa_2!}\left(\frac{x}{\kappa_1(\log x)^{\kappa_1}}\right)^{\kappa_2}}{\frac{1}{(\kappa_2-1)!x^{\kappa_2-1}}} = \frac{x}{\kappa_2\kappa_1^{\kappa_2}(\log x)^{\kappa_1\kappa_2}}. \tag{15}$$

Our theorem now follows since $|C| \leqq |A+B+C|$.

4. Statements and proofs of some technical lemmas

We will need the following three sieve lemmas, and their various corollaries: the Large Sieve of Montgomery (see [12]), Brun's Upper Bound Sieve (see [8]), and a variant of Gallagher's Sieve (see [7]):

LEMMA 4 (Montgomery's Sieve). *Given a set of integers $J \subseteqq \{1,2,\dots,x\}$, and for each prime $p \leqq x$, let $\omega(p)$ be the number of progressions modulo p which J fails to occupy. Then,*

$$|J| \leqq \frac{x+Q^2}{\sum\limits_{q \leqq Q} \mu^2(q) \prod\limits_{p|q} \frac{\omega(p)}{p-\omega(p)}}.$$

One has the following corollary, which essentially appears in Vaughan's paper [16].

Corollary 3. *For J and $\omega(p)$ as above, and $T \leqq \sqrt{x}$, we have*

$$|J| \leqq \frac{2x}{\left(\frac{1}{m}\sum\limits_{p \leqq T} \frac{\omega(p)}{p}\right)^m},$$

where $m = \lfloor (\log x)/(2\log T) \rfloor$.

Note. In Vaughan [16] he proves this result with the factor 4 on the right hand side, instead of the factor 2. The reason is that he used an earlier, weaker form of the Large Sieve.

Lemma 5 (Brun's Sieve). *Suppose that $J \subseteqq \{1, 2, \dots, x\}$ is the largest such set of integers which fails to occupy $\omega(p) \leqq B$ progressions modulo p, for each prime $p \leqq z$. Then,*

$$|J| \ll_B x \prod_{p \leqq z} \left(1 - \frac{\omega(p)}{p}\right).$$

Lemma 6 (Gallagher's Sieve, see [2] and [7]). *Suppose that $J \subseteqq \{1, 2, \dots, x\}$, and $|J| > U$. Then,*

$$|J|^2\bigl(\log x + O(1)\bigr) > \sum_{\substack{p \leqq U \\ p \text{ prime}}} (\log p) \sum_{c=0}^{p-1} \bigl|J(c,p)\bigr|^2.$$

A corollary of this sieve which we will need is the following:

Corollary 4. *Suppose J is as in Lemma 6, and let $h(p)$ denote the number of residue classes modulo p occupied by J, for each $p \leqq U < |J|$. Then,*

$$\log x + O(1) > \frac{1}{|J|^2} \sum_{p \leqq U} (\log p) \sum_{c=0}^{p-1} \bigl|J(c,p)\bigr|^2 \geqq \sum_{p \leqq U} \frac{\log p}{h(p)}.$$

We will also need the following inequality of Cauchy and Davenport (see [13]):

LEMMA 7 (Cauchy–Davenport Inequality). *For sets G and H, let h_1, h_2 and h_3 denote the number of residue classes modulo p occupied by G, H and $G+H$, respectively. Then,*

$$h_3 \geqq \min\left(h_1+h_2-1,p\right).$$

Finally, we will also need the following simple consequence of the Cauchy–Schwarz inequality:

LEMMA 8. *Suppose that J is a set of integers which occupies at most k progressions modulo m. Then,*

$$\sum_{a=0}^{m-1} \left| J(a,m) \right|^2 \geqq \frac{|J|^2}{k}.$$

To prove this lemma, let $\delta(a)$ be 1 if a is in one of the progressions occupied by J (there are at most k such progressions), and let it be 0 otherwise. Then, the lemma follows quickly from the Cauchy–Schwarz inequality:

$$k\sum_{a=0}^{m-1} \left| J(a,m) \right|^2 \geqq \left(\sum_{a=0}^{m-1} \left| J(a,m) \right|^2 \right) \left(\sum_{a=0}^{m-1} \delta(a)^2 \right)$$

$$\geqq \left(\sum_{a=0}^{m-1} \left| J(a,m) \right| \delta(a) \right)^2 = |J|^2.$$

We now prove those of the above lemmas which cannot be found in the literature, as well as Lemmas 1 and 2.

PROOF OF LEMMA 1. To prove this lemma we let $\varepsilon = 1/2$, and let $D>0$ and S be as in the definition of regular sets. Further, let

$$T = \left\{ n \in A_1+\dots+A_k : n \leqq x,\ r(n;A_1,\dots,A_k) \leqq \log^D x \right\}.$$

Then,

$$A_1(x)\cdots A_k(x) = \sum_{n\in T} r(n;A_1,\dots,A_k) + \sum_{n \in S} r(n;A_1,\dots,A_k)$$

$$< (\log^D x)T(x) + \frac{1}{2}A_1(x)\cdots A_k(x)$$

$$\leqq (\log^D x)(A_1+\dots+A_k)(x) + \frac{1}{2}A_1(x)\cdots A_k(x).$$

Rearranging terms gives

$$A_1(x)\cdots A_k(x) < 2(\log^D x)(A_1+\cdots+A_k)(x);$$

and so, the conclusion of the lemma holds with $E = D + 1$.

PROOF OF LEMMA 2. We note that if n satisfies (13) then the largest prime divisor of each of the numbers $n+k, n+2k, \dots, n+rk$ is $< \log^{cr^2} x$. Thus, for each prime $p \in \big[\log^{cr^2} x, x\big)$, we must have that

$$n \not\equiv -k, -2k, \dots, -rk \pmod{p}.$$

Thus, the number of progressions which n can lie in modulo p, for each such p, is $h(p) < p - r$. From Brun's Sieve, we get that the number of integers n satisfying (13) is

$$\ll_r x \prod_{\substack{\log^{cr^2} x < p < Q \\ p \text{ prime}}} \left(1 - \frac{r}{p}\right) < x \exp\left(-r \sum_{\substack{\log^{cr^2} x < p < Q \\ p \text{ prime}}} \frac{1}{p}\right) \tag{16}$$

$$= x \exp\big(-r\big(\log\log x - O(\log\log\log x)\big)\big) = o\left(\frac{x}{\log^{r/2} x}\right),$$

which proves the lemma.

PROOF OF LEMMA 6. We have that for any pair of integers $j_1, j_2 \in J$, $|j_1 - j_2| < x$, and so

$$\sum_{\substack{p \mid j_1 - j_2 \\ p \text{ prime}}} \log p \leqq \log|j_1 - j_2| < \log x.$$

Summing over all pairs j_1, j_2 (of which there are at most $|J|^2$), we get

$$|J|^2 \log x > \sum_{\substack{j_1, j_2 \in J \\ j_1 \neq j_2}} \sum_{\substack{p \mid j_1 - j_2 \\ p \text{ prime}}} \log p$$

$$> \sum_{p \leqq U} (\log p) \sum_{c=0}^{p-1} \#\big\{ j_1, j_2 \in J : j_1 \neq j_2,\ j_1 \equiv j_2 \equiv c \pmod{p}\big\}$$

$$= \sum_{p \leqq U} \left((\log p) \Big(\sum_{c=0}^{p-1} \big| J(c,p) \big|^2 \Big) - |J| \log p \right).$$

Using the fact that

$$\sum_{p \leqq U} \log p = O(U) = O\big(|J| \big),$$

and rearranging terms in the above string of inequalities, we get

$$|J|^2 \big(\log x + O(1) \big) > \sum_{p \leqq U} (\log p) \sum_{c=0}^{p-1} \big| J(c,p) \big|^2,$$

as claimed.

Proof of Corollary 4. Since J occupies $h(p)$ progressions modulo p, we have from Lemma 8 that

$$\frac{1}{|J|^2} \sum_{c=0}^{p-1} \big| J(c,p) \big|^2 > \frac{1}{h(p)}.$$

Putting this into Lemma 6, we get

$$\log x + O(1)) > \frac{1}{|J|^2} \sum_{\substack{p \leqq U \\ p \text{ prime}}} (\log p) \sum_{c=0}^{p-1} \big| J(c,p) \big|^2 > \sum_{p \leqq U} \frac{\log p}{h(p)},$$

as claimed.

5. Proof of Proposition 1

Let $\tau = \lfloor \delta \rfloor + 1$. Then, we have $|G| \geqq |F| \geqq \tau$.

The proof involves four iterations: In the first iteration we will show that $|G| \ll x(\log x)^{-\tau + o(1)}$, and thus $|F| \gg (\log x)^{\tau - \delta - o(1)}$; in the second iteration, we will show that $|F| \gg \exp\big((\log x)^{\tau - \delta} \big)$; in the third iteration, we will show that $|F| > x^{1/3}$, for x sufficiently large; and, in the final iteration, we will show that

$$\frac{x^{1/2}}{\log^{\delta+6}} \ll |F| \leqq |G| \ll x^{1/2} \log^6 x.$$

We note that our proposition can be proved using three iterations (instead of four), as was done in [4]; also, no attempt was made to optimize the powers of the logarithms appearing in the result.

Throughout the proof we let $h_1(p)$ and $h_2(p)$ denote the number of residue classes occupied by F and G, respectively. Since no element of $F+G$ can be divisible by a prime $\leqq \sqrt{x}$, we deduce that $F+G$ occupies at most $p-1$ residue classes modulo p for each such prime. So, from Lemma 7, we deduce

$$h_1(p) + h_2(p) \leqq p. \tag{17}$$

We let $\omega(p) = p - h_2(p)$ be the number of progressions which G fails to occupy; and so, (17) implies that $\omega(p) \geqq h_1(p)$.

For the first iteration, let $f_1, \ldots, f_\tau$ be any τ elements of F, and let Z be the set of primes $\leqq \sqrt{x}$ with the property that $f_1, \ldots, f_\tau$ all occupy different residue classes modulo p. Let P be the set of primes $\leqq \sqrt{x}$, and set

$$f(Z) = \sum_{p \in Z} \frac{1}{p} = \log\log x + O(1) - \sum_{p \in P \setminus Z} \frac{1}{p}.$$

To estimate this last sum, we first define

$$s(n) = \sum_{\substack{p|n \\ p \text{ prime}}} \frac{1}{p}.$$

Then, $s(n) \ll \log\log\log n$, and this upper bound is attained when n is the product of the primes $\leqq \log n$. Now, if $p \in P \setminus Z$, then $p|\Delta$, where

$$\Delta = \prod_{1 \leqq i < j \leqq \tau} |f_j - f_i| \ll x^{\tau^2/2};$$

and so,

$$\sum_{p \in P \setminus Z} \frac{1}{p} \leqq \sum_{p|\Delta} \frac{1}{p} \ll_\tau \log\log\log x.$$

Thus,

$$f(Z) = \log\log x - O_\tau(\log\log\log x).$$

Letting $\Pi(x)$ be the product of the primes $\leqq \sqrt{x}$, we deduce from Lemma 5 that

$$|G| \leqq \#\big\{n \leqq x : \big((n+f_1)(n+f_2)\cdots(n+f_\tau),\Pi(x)\big) = 1\big\}$$
$$\ll_\tau x \prod_{p\in Z}\left(1-\frac{\tau}{p}\right) \ll x\exp\left(-\tau\sum_{p\in Z}\frac{1}{p}\right) < \frac{x}{\log^{\tau-o(1)} x}.$$

Thus, since $x(\log x)^{-\delta} \ll |F|\,|G|$, we deduce $|F| \gg (\log x)^{\tau-\delta-o(1)}$, as claimed.

For the second iteration, let $f_1,\dots,f_t \in F$, where $t = \log^{\tau-\delta-o(1)} x$, and, as before, let Z' be the set of primes $\leqq \sqrt{x}$ where all the f_i's fall into distinct residue classes modulo p. Then, as before, let

$$\Delta' = \prod_{1\leqq i<j\leqq t} |f_i - f_j| \ll x^{t^2/2}.$$

Then, $f_1,\dots,f_t$ are not distinct modulo p implies $p|\Delta'$. As before, we have

$$\sum_{p\in P\setminus Z'}\frac{1}{p} \leqq \sum_{p|\Delta'}\frac{1}{p} \ll \log\log\log x.$$

Thus, if we let $T = \exp\left(\log^{1-\tau/2+\delta/2} x\right)$, then

$$\sum_{\substack{p\leqq T\\ p\in Z'}}\frac{1}{p} \geqq \sum_{\substack{p\leqq T\\ p \text{ prime}}}\frac{1}{p} - \sum_{p\in P\setminus Z'}\frac{1}{p} = \log\log T - O(\log\log\log x).$$

Now, applying Corollary 3 with

$$m = \left\lfloor (\log^{\tau/2-\delta/2} x)/2 \right\rfloor \quad\text{and}\quad \omega(p) = (\log x)^{\tau-\delta-o(1)} \quad\text{for all}\quad p\in Z'\cap[2,T],$$

we get

$$|G| \leqq \#\big\{n \leqq x : \big((n+f_1)(n+f_2)\cdots(n+f_t),\Pi(x)\big) = 1\big\}$$
$$\leqq \frac{x}{\Big((\log x)^{\tau/2-\delta/2-o(1)}\sum_{\substack{p\leqq T\\ p\in Z'}}\frac{1}{p}\Big)^m} \ll \frac{x}{\exp(2m)} \ll \frac{x}{\exp\left(\log^{\tau/2-\delta/2} x\right)}.$$

Thus, since $x/\log^\delta x < |F|\,|G|$, we conclude that

$$|F| > \exp\Big(\big(1-o(1)\big)\sqrt{\log^{\tau-\delta} x}\Big).$$

For the third iteration, let

$$T' = \exp\left(\sqrt{\log^{\tau-\delta} x}/2\right) \quad \text{and} \quad m' = \left\lfloor \sqrt{\log^{2-\tau+\delta} x} \right\rfloor.$$

Then, from Corollary 4, we have

$$\log x + O(1) > \sum_{T'/2 \leqq p \leqq T'} \frac{\log p}{h_1(p)}.$$

(Note that we use the corollary with $J = F$, and we have from iteration two that $|F| > T'$ for x sufficiently large.) So, for almost all primes $p \in [T'/2, T']$ we have that $\omega(p) \geqq h_1(p) > p/\log^2 x$; and so,

$$\sum_{T'/2 \leqq p \leqq T} \frac{\omega(p)}{p} \gg \frac{T'}{\log^3 x}.$$

Using Corollary 3 with $T = T'$, we deduce

$$|G| \ll \frac{x}{\left(\frac{1}{m'} \sum_{T'/2 \leqq p \leqq T'} \frac{\omega(p)}{p}\right)^{m'}} < \frac{x}{x^{1/2-o(1)}} = x^{1/2+o(1)}.$$

Thus, since $|F|\,|G| \gg x \log^{-\delta} x$, we deduce $|F| > x^{1/3}$ for x sufficiently large.

For the last iteration, we have by Corollary 4 that

$$\log x + O(1) > \sum_{x^{1/4}/2 \leqq p \leqq x^{1/4}} \frac{\log p}{h_1(p)};$$

and it follows that almost all primes in $\left[x^{1/4}/2, x^{1/4}\right]$ have $\omega(p) \geqq h_1(p) > p/\log^2 x$. Thus,

$$\sum_{x^{1/4}/2 \leqq p \leqq x^{1/4}} \frac{\omega(p)}{p} \gg \frac{x^{1/4}}{\log^3 x}.$$

By Corollary 3 we have

$$|G| \ll \frac{x}{\left(\sum_{x^{1/4}/2 \leqq p \leqq x^{1/4}} \frac{\omega(p)}{p}\right)^2} \ll \frac{x \log^6 x}{x^{1/2}} = \sqrt{x} \log^6 x;$$

and so, since $x/\log^{\delta} x \ll |F|\,|G|$, we deduce

$$\frac{\sqrt{x}}{\log^{6+\delta} x} \ll |F| \leqq |G| \ll \sqrt{x}\log^{6} x,$$

and the proposition is proved.

6. Proof of Proposition 2

Since A, B, C is a regular triple of sets, we have from Lemma 1 that for some $E > 0$,

$$\frac{A\big(\sqrt{x}\,\big)}{A(x)} \cdot \frac{B\big(\sqrt{x}\,\big)}{B(x)} \cdot \frac{C\big(\sqrt{x}\,\big)}{C(x)} \leqq (\log^{E} x)\frac{(A+B+C)\big(\sqrt{x}\,\big)}{(A+B+C)(x)} \tag{18}$$

$$\ll \frac{\sqrt{x}\log^{E} x}{x/\log^{\kappa} x} = \frac{\log^{E+\kappa} x}{\sqrt{x}}.$$

Thus, for x sufficiently large, one of the following inequalities must hold:

$$A\big(\sqrt{x}\,\big) < \frac{A(x)}{x^{1/5}} \quad \text{or} \quad B\big(\sqrt{x}\,\big) < \frac{B(x)}{x^{1/5}} \quad \text{or} \quad C\big(\sqrt{x}\,\big) < \frac{C(x)}{x^{1/5}}. \tag{19}$$

Suppose that the inequality holds for $A(x)$ and $A\big(\sqrt{x}\,\big)$. Then, letting

$$\hat{A} = A \cap \big(\sqrt{x}, x\big], \quad \hat{B} = B \cap [1,x], \quad \text{and} \quad \hat{C} = C \cap [1,x]$$

gives

$$|\hat{A}| \sim A(x), \quad |\hat{B}| \sim B(x), \quad |\hat{C}| \sim C(x), \quad \text{and} \quad |\hat{A}+\hat{B}+\hat{C}| \subset \big(\sqrt{x}, x\big].$$

Also, since A, B, C is a regular triple, we get

$$0 \leqq (A+B+C)(x) - |\hat{A}+\hat{B}+\hat{C}|$$

$$\leqq \#\big\{n = a+b+c :\ a \in A, b \in B, c \in C,\ a \leqq \sqrt{x}\big\}$$

$$\leqq A\big(\sqrt{x}\,\big)B(x)C(x) \leqq x^{-1/5}A(x)B(x)C(x)$$

$$\leqq x^{-1/5}(\log^{E} x)(A+B+C)(x).$$

Thus,

$$(A+B+C)(x) \sim |\hat{A}+\hat{B}+\hat{C}|,$$

as claimed. We get the same conclusions for the remaining cases of (19).

7. Proof of Proposition 3

Since A, B, C is a regular triple, one can easily deduce that for $\varepsilon = 1/12$ and x sufficiently large, there exists $E > 0$ such that if $|S| \leqq \kappa$, then

$$\sum_{\substack{n \in L\times\hat{C} \\ r(n;L,\hat{C}) > \log^E x}} r(n;L,\hat{C}) < \varepsilon|L\times\hat{C}|. \tag{20}$$

For the remainder of the proof of this proposition, we will assume that E is such that this inequality is satisfied.

The proof now proceeds using a probabilistic argument: Let L' and C' be random subsets of L and $\hat{C}$, respectively, where

$$\operatorname{Prob}(\ell \in L' \mid \ell \in L) = \operatorname{Prob}(c \in C' \mid c \in \hat{C}) = \frac{1}{\log^{2E} x},$$

where all these probabilities are independent. Clearly, $|L'|$ and $|C'|$ each have a binomial distribution, which implies that the following occurs with probability $1 - o(1)$:

$$\frac{E\big(|L'\times C'|\big)}{2} < |L'\times C'| < 2E\big(|L'\times C'|\big), \tag{21}$$

where $E\big(|L'\times C'|\big)$ is the usual expectation given by

$$E\big(|L'\times C'|\big) = \sum_{(\ell,c)\in L\times\hat{C}} \operatorname{Prob}\big((\ell,c)\in L'\times C'\big) = \frac{|L\times\hat{C}|}{\log^{4E} x}.$$

In the course of our proof, we will show that the event

$$(1-6\varepsilon)|L'\times C'| < |L'+C'| \quad \text{and (21) occurs} \tag{22}$$

has positive probability, which will imply that there exist subsets $L^* \subset L$ and $C^* \subset \hat{C}$ satisfying these same inequalities. If we can do this, then (12) will hold (since $\varepsilon = 1/12$), and (11) will hold for $D = E + 1$ and x sufficiently large.

Thus, the proposition will follow if we can show that (22) has positive probability. We note that it suffices to prove that

$$\text{Prob}\left(|L' \times C'| - |L' + C'| < 3\varepsilon E(L' \times C')\right) > \frac{1}{2}, \tag{23}$$

since (21) holds with probability $1 - o(1)$.

We now proceed to show that (23) holds: Suppose that $n \in L + \hat{C}$ has exactly k solutions to

$$n = \ell_1 + c_1, \dots, \ell_k + c_k, \quad \text{each} \quad (\ell_i, c_i) \in L \times \hat{C}.$$

Then, since the ℓ_i's are distinct, and the c_i's distinct, we have that all subsets of the probabilities

$$\text{Prob}\left((\ell_1, c_1) \in L' \times C'\right), \dots, \text{Prob}\left((\ell_k, c_k) \in L' \times C'\right) = \frac{1}{\log^{4E} x}$$

are independent. It follows then that if we let $r'(n)$ be the random variable

$$r'(n) = \left\{(\ell, c) \in L' \times C' : n = \ell + c\right\},$$

then $\text{Prob}\left(r'(n) = d\right)$ has a binomial distribution, given by

$$\text{Prob}\left(r'(n) = d\right) = \binom{k}{d}\left(1 - \frac{1}{\log^{4E} x}\right)^{k-d} \frac{1}{\log^{4dE} x} < \frac{k^d}{d! \log^{4dE} x};$$

and, we have the easily checked expectation formula

$$E\left(r'(n)\right) = \frac{r(n; L, \hat{C})}{\log^{4E} x},$$

where $r(n; L, \hat{C})$ is as defined in the Introduction.

For bookkeeping purposes, define

$$N = \left\{n \in L + \hat{C} : n \text{ has at most } \log^E x \text{ solutions to}\right.$$
$$\left. = \ell + c,\ (\ell, c) \in L \times \hat{C}\right\};$$

and $\overline{N} = (L + \hat{C}) \setminus N$, and define the random variable

$$\delta(n) = \begin{cases} 0, & \text{if } \; n \notin L' + C'; \\ 1, & \text{if } \; n \in L' + C'. \end{cases}$$

Then, from (20) and the above probability and expectation estimates, we have:

$$E\big(|L' \times C'| - |L' + C'|\big) = \sum_{n \in L + \hat{C}} E\big(r'(n) - \delta(n)\big)$$

$$= \sum_{n \in N} E\big(r'(n) - \delta(n)\big) + \sum_{n \in \overline{N}} E\big(r'(n) - \delta(n)\big)$$

$$\leqq \sum_{n \in N} \sum_{d \geqq 2} (d-1) \operatorname{Prob}\big(r'(n) = d\big) + \sum_{n \in \overline{N}} E\big(r'(n)\big)$$

$$\leqq \sum_{n \in N} \sum_{d \geqq 2} \frac{1}{(d-1)! \log^{3dE} x} + \frac{1}{\log^{4E} x} \sum_{n \in \overline{N}} r(n; L, \hat{C})$$

$$\leqq \frac{2|L \times \hat{C}|}{\log^{6E} x} + \frac{\varepsilon |L \times \hat{C}|}{\log^{4E} x} = E\big(|L' \times C'|\big) \left(\varepsilon + \frac{2}{\log^{2E} x}\right).$$

Let us recall Markov's Inequality. If X is a non-negative random variable, then

$$\operatorname{Prob}(X \geqq a) \leqq \frac{E(X)}{a}.$$

From this inequality with $X = |L' \times C'| - |L' + C'|$, together with our above expectation estimates, we deduce

$$\operatorname{Prob}\big(|L' \times C'| - |L' + C'| \geqq 3\varepsilon E\big(|L' \times C'|\big)\big) < \frac{1}{3} - \frac{3}{\log^{4E} x}.$$

Therefore, (23) holds for x sufficiently large.

8. Proof of Proposition 4, Corollary 1, and Lemma 9

For a given set of integers J, let J_p denote the set of residue classes modulo p occupied by J, and let $\overline{J_p}$ denote those residue classes *not* occupied by J. Clearly, $|\overline{J_p}| = p - |J_p|$.

For all integers j we have that $(\ell, c) \in L^{\#} \times C^{*}$ is a solution to $\ell + c + jk \equiv 0 \pmod{p}$ if and only if

$$(\ell, c) \equiv (r, -r - jk) \pmod{p}, \quad \text{for some} \quad r \in L^{\#}_p \setminus (L^{\#} - jk)_p.$$

From this and Cauchy's inequality we have

$$Z := \sum_{p \leqq Q} (\log p) \# \big\{ (\ell, c) \in L^{\#} \times C^{*} : \ell + c + jk \equiv 0 \pmod{p} \big\}$$

$$= \sum_{p \leqq Q} (\log p) \sum_{r \in L^{\#}_p \setminus (L^{\#} - jk)_p} \big| L^{\#}(r, p) \big| \, \big| C^{*}(-r - jk, p) \big| \leqq Z_1^{1/2} Z_2^{1/2},$$

where

$$Z_1 = \sum_{p \leqq Q} (\log p) \sum_{r \in L^{\#}_p \setminus (L^{\#} - jk)_p} \big| L^{\#}(r, p) \big|^2;$$

and

$$Z_2 = \sum_{p \leqq Q} (\log p) \sum_{r \in L^{\#}_p \setminus (L^{\#} - jk)_p} \big| C^{*}(-r - jk, p) \big|^2.$$

To bound Z_1 and Z_2 from above we will require the following three results:

Lemma 9. *We have for $Q = \sqrt{x} \log^{O(1)} x$ that*

$$\sum_{p \leqq Q} \frac{\log p}{|L^{\#}_p|} = \log x + O(\log \log x) = \sum_{p \leqq Q} \frac{\log p}{|C^{*}_p|}; \tag{24}$$

and

$$\sum_{p \leqq Q} \frac{\log p}{p - |L^{\#}_p|} = \log x + O(\log \log x) = \sum_{p \leqq Q} \frac{\log p}{p - |C^{*}_p|}. \tag{25}$$

For similar results see Elsholtz [5].

PROPOSITION 5. *Suppose that $J = L^{\#}$ or C^*, and let K be the other set (if $J = L^{\#}$, then $K = C^*$, and vice versa). Also, suppose that for each prime $p \leqq Q$ we have a set of residue classes $G_p \subseteqq J_p$. Then, we have the following inequality:*

$$\sum_{p \leqq Q} (\log p) \sum_{r \in G_p} \bigl| J(r,p) \bigr|^2 < |J|^2 \left(\sum_{p \leqq Q} \frac{(\log p)|G_p|}{\bigl(p - |K_p|\bigr)^2} + O(\log\log x) \right).$$

LEMMA 10. *Suppose $J = L^{\#}$ or C^*, and that K is the other of the two sets. Then, for any integer $j > 0$,*

$$\sum_{p \leqq Q} \frac{(\log p) \bigl| J_p \setminus (J - jk)_p \bigr|}{p - |K_p|^2} = O(j \log\log x).$$

The proofs of these last two results will make use of the following basic facts about the sets $L^{\#}$ and C^*: Since for every $(\ell, c) \in L^{\#} \times C^*$ we have $\ell + c$ and $\ell + c + k$ are primes $> \sqrt{x}$, there can be no solutions to $\ell + c \equiv 0 \pmod{p}$ or $\ell + c + k \equiv 0 \pmod{p}$ for any prime $p \leqq Q < \sqrt{x}$. Thus,

$$\text{(26)} \qquad \begin{cases} L_p^{\#} \cap (-C^*)_p = \emptyset = (L^{\#} + k)_p \cap (-C^*)_p \\ \Longrightarrow L_p^{\#} \quad \text{and} \quad (L^{\#} + k)_p \quad \text{are both subsets of} \quad \overline{(-C^*)_p}. \end{cases}$$

Similarly,

$$\text{(27)} \qquad C_p^* \quad \text{and} \quad (C^* + k)_p \quad \text{are both subsets of} \quad \overline{(-L^{\#})_p}.$$

Resuming the proof of our Proposition 4, we have from Proposition 5 and Lemma 10 with $J = L^{\#}$ and $G_p = L_p^{\#} \setminus (L^{\#} - jk)_p$ that

$$Z_1 < |L^{\#}|^2 \left(\sum_{p \leqq Q} \frac{(\log p)|G_p|}{\bigl(p - |K_p|\bigr)^2} + O(\log\log x) \right) = O(j |L^{\#}|^2 \log\log x).$$

Applying these two results with $J = C^*$ and

$$G_p = \bigl(- \bigl(L^{\#} \setminus (L^{\#} - jk)_p \bigr) - jk \bigr)_p$$

(note: $|G_p| = \bigl| L^{\#} \setminus (L^{\#} - jk)_p \bigr|$), we likewise get

$$Z_2 = O\bigl(j |C^*|^2 \log\log x \bigr).$$

Thus,

$$Z \leqq Z_1^{1/2} Z_2^{1/2} = O\bigl(j|L^{\#}|\,|C^*| \log\log x\bigr),$$

which proves the proposition.

Proof of Corollary 1. We have that

$$\sum_{n \in L^{\#}+C^*} \left(\sum_{j=1}^{r} \sum_{\substack{p \leqq Q \\ p \mid n+jk \\ p \text{ prime}}} \log p \right) \leqq \sum_{(\ell,c) \in L^{\#} \times C^*} \sum_{j=1}^{r} \sum_{\substack{p \leqq Q \\ p \mid \ell+c+jk \\ p \text{ prime}}} \log p$$

$$= \sum_{j=1}^{r} \sum_{p \leqq Q} \#\bigl\{ (\ell, c) \in L^{\#} \times C^* : p \mid \ell + c + jk \bigr\}$$

$$= \sum_{j=1}^{r} O\bigl(j|L^{\#}|\,|C^*| \log\log x\bigr) = O\bigl(r^2 |L^{\#} + C^*| \log\log x\bigr),$$

where this last equality follows from Proposition 3. It is now obvious that more than half the elements $n \in L^{\#} + C^*$ satisfy (13), which proves the corollary.

Proof of Lemma 9. Since for $p \leqq Q$ the set $L^{\#} + C^*$ contains no numbers $\equiv 0 \pmod{p}$, it follows from Lemma 7 that $|L_p^{\#}| + |C_p^*| \leqq p$; and so,

$$\frac{1}{|L_p^{\#}|} + \frac{1}{|C_p^*|} \geqq \frac{1}{|L_p^{\#}|} + \frac{1}{p - |L_p^{\#}|} \geqq \frac{4}{p}.$$

From this inequality we deduce

$$\text{(28)} \qquad \sum_{p \leqq Q} (\log p) \left(\frac{1}{|L_p^{\#}|} + \frac{1}{|C_p^*|} \right) \geqq \sum_{p \leqq Q} (\log p) \left(\frac{1}{|L_p^{\#}|} + \frac{1}{p - |L_p^{\#}|} \right)$$

$$\geqq \sum_{p \leqq Q} \frac{4 \log p}{p} = 2 \log x + O(\log\log x).$$

Now, from Corollary 4 with $J = L^{\#}$ and $J = C^*$ we deduce

$$\sum_{p \leqq Q} \frac{\log p}{|L_p^{\#}|} < \log x + O(\log\log x) \quad \text{and} \quad \sum_{P \leqq Q} \frac{\log p}{|C_p^*|} < \log x + O(\log\log x).$$

Combining these two upper bounds with (28), we have that (24) is satisfied; and

$$\sum_{p \leqq Q} (\log p) \left(\frac{1}{|L_p^{\#}|} + \frac{1}{p - |L_p^{\#}|} \right) = 2 \log x + O(\log \log x).$$

This equation and (24) together imply that

$$\sum_{p \leqq Q} \frac{\log p}{p - |L_p^{\#}|} = \sum_{p \leqq Q} (\log p) \left(\frac{1}{|L_p^{\#}|} + \frac{1}{p - |L_p^{\#}|} \right) - \sum_{p \leqq Q} \frac{\log p}{|L_p^{\#}|}$$

$$= \log x + O(\log \log x),$$

which gives that the first part of (25) is satisfied. The second part of (25) is satisfied by applying the same argument.

9. Proof of Proposition 5

Let

$$V_p(r) = \left(J(r, p) - \frac{|J|}{p - |K_p|} \right)^2.$$

The sum we wish to bound from above is as follows:

$$X := \sum_{p \leqq Q} (\log p) \sum_{r \in G_p} J(r, p)^2 \tag{29}$$

$$\leqq \sum_{p \leqq Q} (\log p) \sum_{r \in G_p} \left(V_p(r) - 2 \frac{J(r, p)|J|}{p - |J_p|} + \frac{|J|^2}{\bigl(p - |J_p|\bigr)^2} \right)$$

$$< \sum_{p \leqq Q} (\log p) \sum_{r \in J_p} V_p(r) + |J|^2 \sum_{p \leqq Q} (\log p) \sum_{r \in G_p} \frac{1}{\bigl(p - |J_p|\bigr)^2}.$$

Now, we have that

$$Y := \sum_{p \leqq Q} (\log p) \sum_{r \in \overline{(-K)_p}} V(r, p) = E_1 - 2E_2 + E_3,$$

where

$$E_1 = \sum_{p \leqq Q} (\log p) \sum_{r \in \overline{(-K)}_p} J(r,p)^2$$

$$= \sum_{p \leqq Q} (\log p) \sum_{r \in J_p} J(r,p)^2 = |J|^2 \bigl(\log x + O(1) \bigr),$$

$$E_2 = \sum_{p \leqq Q} (\log p) \sum_{r \in \overline{(-K)}_p} \frac{J(r,p)|J|}{p - |J_p|} = \sum_{p \leqq Q} (\log p) \sum_{r \in J_p} \frac{J(r,p)|J|}{p - |J_p|}$$

$$= |J|^2 \sum_{p \leqq Q} \frac{\log p}{|p - K_p|} = |J|^2 \bigl(\log x + O(\log\log x) \bigr),$$

$$E_3 = |J|^2 \sum_{p \leqq Q} \frac{\log p}{p - |J_p|} = E_2.$$

Note that the upper bound we derived for E_1 comes from Lemma 6, together with the fact that $\log Q = (\log x)/2 + O(\log\log x)$; the equation for E_2 comes from Lemma 9; and, the switching of some of the above sums from a sum over $r \in \overline{(-K)}_p$ to $r \in J_p$ is justified since $J_p \subseteqq \overline{(-K)}_p$, by (26) and (27).

It follows that

$$Y = E_1 - E_2 = O\bigl(|J|^2 \log\log x \bigr).$$

Substituting this into (29) gives

$$X < Y + |J|^2 \sum_{p \leqq Q} (\log p) \sum_{r \in G_p} \frac{1}{\bigl(p - |J_p|\bigr)^2}$$

$$= |J|^2 \left(\sum_{p \leqq Q} \frac{(\log p)|G_p|}{\bigl(p - |J_p|\bigr)^2} + O(\log\log x) \right),$$

which proves the proposition.

10. Proof of Lemma 10

For an integer h, let $S(h)$ denote the symmetric difference between $(J-hk)_p$ and $\bigl(J-(h-1)k\bigr)_p$. We note that $\bigl|S(h)\bigr| = \bigl|S(0)\bigr|$.

Now, since

$$J_p \setminus (J-hk)_p \subseteqq S(h) \cup \Bigl(J_p \setminus \bigl(J-(h-1)k\bigr)_p\Bigr),$$

it follows that

$$\bigl|J_p \setminus (J-hk)_p\bigr| \leqq \bigl|S(0)\bigr| + \bigl|J_p \setminus \bigl(J-(h-1)k\bigr)_p\bigr|.$$

For $h \geqq 1$ a simple induction argument then shows that

$$\bigl|J_p \setminus (J-hk)_p\bigr| \leqq h\bigl|S(0)\bigr|. \tag{30}$$

Now, from (26) and (27) we deduce that $J_p, (J+k)_p \subseteqq \overline{(-K)_p}$, which gives:

$$\begin{aligned}\bigl|J_p \setminus (J+k)_p\bigr| &\leqq \bigl|\overline{(-K)_p} \setminus (J+k)_p\bigr| \\ = \bigl|\overline{(-K)_p}\bigr| - \bigl|(J+k)_p\bigr| &= p - |K_p| - |J_p|;\end{aligned}$$

and

$$\bigl|(J+k)_p \setminus J_p\bigr| \leqq \bigl|\overline{(-K)_p} \setminus J_p\bigr| = p - |K_p| - |J_p|.$$

Thus,

$$\bigl|S(0)\bigr| = \bigl|J_p \setminus (J+k)_p\bigr| + \bigl|(J+k)_p \setminus J_p\bigr| \leqq 2\bigl(p - |K_p| - |J_p|\bigr). \tag{31}$$

From this and the fact that

$$p - |K_p| = \bigl|\overline{(-K)_p}\bigr| \leqq |J_p|,$$

we deduce

$$\begin{aligned}\sum_{p\leqq Q} (\log p)\frac{\bigl(p-|K_p|-|J_p|\bigr)}{\bigl(p-|K_p|\bigr)^2} &\leqq \sum_{p\leqq Q} (\log p)\frac{\bigl(p-|K_p|-|J_p|\bigr)}{|J_p|\bigl(p-|K_p|\bigr)} \\ = \sum_{p\leqq Q} (\log p)\left(\frac{1}{|J_p|} - \frac{1}{p-|K_p|}\right) &= O(\log\log x),\end{aligned}$$

by Lemma 9. From this, (30), and (31), we deduce

$$\sum_{p\leqq Q}(\log p)\frac{\left|\,J_p\setminus(J-jk)_p\right|}{\bigl(p-|K_p|\bigr)^{\,2}}\leqq 2j\sum_{p\leqq Q}(\log p)\frac{\bigl(p-|K_p|-|K_p|\bigr)}{\bigl(p-|K_p|\bigr)^{\,2}}$$

$$=O(j\log\log x),$$

which proves the lemma.

References

[1] D. Bshouty and N. H. Bshouty, A note on prime n-tuples, *Rocky Mountain J. Math.,* **27** (1997), 775–778.

[2] E. S. Croot III and C. Elsholtz, On variants of the larger sieve, *Acta Math. Hungar.,* **103** (2004), 243–254.

[3] C. Elsholtz, A remark on Hoffman and Wolke's additive decompositions of the set of primes, *Arch. Math,* **76** (2001), 30–33.

[4] C. Elsholtz, The inverse Goldbach problem, *Mathematika,* **48** (2001), 151–158.

[5] C. Elsholtz, Some remarks on the additive structure of the set of primes, *Number Theory for the Millennium* (eds. Bruce Berndt et al.), A. K. Peters (2002), pp. 419–427.

[6] P. Erdős, C. L. Stewart and R. Tijdeman, Some Diophantine equations with many solutions, *Compositio Math.,* **66** (1988), 37–56.

[7] P. X. Gallagher, A larger sieve, *Acta Arith,* **18** (1971), 77–81.

[8] H. Halberstam and H.-E. Richert, *Sieve Methods* (London Mathematical Society Monographs No. 4), Academic Press (1974).

[9] G. H. Hardy and J. E. Littlewood, Some problems of 'Partitio Numerurum', III. On the expression of a number as a sum of primes, *Acta Math.,* **44** (1923), 1–70.

[10] A. Hofmann and D. Wolke, On additive decompositions of the set of primes, *Arch. Math.,* **67** (1996), 379–382.

[11] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen,* Teubner Verlag (Leipzig, Berlin, 1909).

[12] H. Montgomery, The analytic principle of the large sieve, *Bull. Amer. Math Soc,* **84** (1978), 547–567.

[13] M. Nathanson, *Additive Number Theory, Inverse Problems and the Geometry of Sumsets,* Graduate Texts in Mathematics, 165, Springer-Verlag (New York, 1996).

[14] H.-H. Ostmann, *Additive Zahlentheorie,* 1. Teil: Allgemeine Untersuchungen, Springer-Verlag (Berlin–Heidelberg–New York, 1958).

[15] C. Pomerance, C. L. Stewart and A. Sárközy, On divisors of sums of integers, III, *Pacific J. Math.,* **133** (1988), 363–379.

[16] R. C. Vaughan, Some applications of Montgomery's sieve, *J. Number Theory,* **5** (1973), 64–79.

[17] E. Wirsing, *On the additive decomposibility of the set of primes,* unpublished manuscript (Oberwolfach abstracts 28/1972).

(Received June 26, 2003)

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF MATHEMATICS
125 SKILES
ATLANTA, GA 30332
U.S.A.

DEPARTMENT OF MATHEMATICS
ROYAL HOLLOWAY, UNIVERSITY OF LONDON
EGHAM
SURREY TW20 0EX
U.K.