ON THE INFINITUDE OF COUSIN PRIMES

T. AGAMA

ABSTRACT. In this paper we prove that there infinitely many cousin primes by deducing the lower bound

$$\sum_{\substack{p \leq x \\ p, p+4 \in \mathbb{P} \setminus \{2\}}} 1 \geq (1+o(1)) \frac{x}{2\mathcal{C} \log^2 x}$$

where $\mathcal{C} := \mathcal{C}(4) > 0$ fixed and $\mathbb{P}$ is the set of all prime numbers. In particular it follows that

$$\sum_{p, p+4 \in \mathbb{P} \setminus \{2\}} 1 = \infty$$

by taking $x \longrightarrow \infty$ on both sides of the inequality. We start by developing a general method for estimating correlations of the form

$$\sum_{n \leq x} G(n)G(n+l)$$

for a fixed $1 \leq l \leq x$ and where $G : \mathbb{N} \longrightarrow \mathbb{R}^+$.

1. Introduction and statement

The area method developed in [1] serves as universal tool and a black box for studying problem related to correlations. The applications are vast, as it allows us to study the distribution of certain class of integers including but not limited to the primes. The area method, by itself, allows one to decompose any correlated sum of the forms below

$$\sum_{n \leq x} G(n)G(x-n)$$

and

$$\sum_{n \leq x} G(n)G(n+l)$$

where $1 \leq l \leq x$ for some $G : \mathbb{N} \longrightarrow \mathbb{R}$ into double sums under certain local condition that can easily be handled using just classical tools like the partial summation or the Riemann-Stieltjes integration by parts. It turns out this method can also be very much adapted to similar problems like the distribution of sexy primes and the twin prime conjecture, which has been applied and can be found in [3]. In this paper, we prove the infinitude of primes that are distance four apart. The same approach used in [3] is still used in the current paper, except for a change in the shift. In particular, we prove the following result

Date: May 31, 2022.
2010 *Mathematics Subject Classification.* Primary 11Pxx, 11Bxx; Secondary 11Axx, 11Gxx.
Key words and phrases. prime; cousin; area method.

Theorem 1.1. *Let $\mathbb{P}$ denotes the set of all prime numbers, then we have the estimate*

$$\#\{p \leq x \mid p+4, p \in \mathbb{P} \setminus \{2\}\} \geq (1+o(1))\frac{1}{2\mathcal{D}(4)}\frac{x}{\log^2 x}$$

where $\mathcal{D}(4) > 0$ fixed.

In the sequel, for any $f, g : \mathbb{N} \longrightarrow \mathbb{R}$, we will write $f(n) = o(1)$ to mean $\lim\limits_{n \longrightarrow \infty} f(n) = 0$. Also $f(n) \ll g(n)$ would mean there exist some constant $c > 0$ such that $f(n) \leq cg(n)$ for all sufficiently large values of n. The following equivalence $f(n) \sim g(n)$ if and only if $\lim\limits_{n \longrightarrow \infty} \frac{f(n)}{g(n)} = 1$ is also standard notation.

2. The area method

This section introduces and develops a fundamental strategy for solving problems involving arithmetic function correlations. This method is basic in that it employs the attributes of four primary geometric shapes: the triangle, trapezium, rectangle, and square. Exploiting the regions of these forms and putting them together in a coherent manner will result in the basic identity we will generate.

Theorem 2.1. *Let $\{r_j\}_{j=1}^{n}$ and $\{h_j\}_{j=1}^{n}$ be any sequence of real numbers, and let r and h be any real numbers satisfying $\sum\limits_{j=1}^{n} r_j = r$ and $\sum\limits_{j=1}^{n} h_j = h$, and*

$$(r^2 + h^2)^{1/2} = \sum_{j=1}^{n} (r_j^2 + h_j^2)^{1/2},$$

then

$$\sum_{j=2}^{n} r_j h_j = \sum_{j=2}^{n} h_j \bigg(\sum_{i=1}^{j} r_i + \sum_{i=1}^{j-1} r_i \bigg) - 2\sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}.$$

Proof. Consider a right angled triangle, say ΔABC in a plane, with height h and base r. Next, let us partition the height of the triangle into n parts, not necessarily equal. Now, we link those partitions along the height to the hypotenuse, with the aid of a parallel line. At the point of contact of each line to the hypotenuse, we drop down a vertical line to the next line connecting the last point of the previous partition, thereby forming another right-angled triangle, say $\Delta A_1B_1C_1$ with base and height r_1 and h_1 respectively. We remark that this triangle is covered by the triangle ΔABC, with hypotenuse constituting a proportion of the hypotenuse of triangle ΔABC. We continue this process until we obtain n right-angled triangles $\Delta A_jB_jC_j$, each with base and height r_j and h_j for $j = 1, 2, \ldots n$. This construction satisfies

$$h = \sum_{j=1}^{n} h_j \text{ and } r = \sum_{j=1}^{n} r_j$$

and

$$(r^2 + h^2)^{1/2} = \sum_{j=1}^{n} (r_j^2 + h_j^2)^{1/2}.$$

Now, let us deform the original triangle ΔABC by removing the smaller triangles $\Delta A_jB_jC_j$ for $j = 1, 2, \ldots n$. Essentially we are left with rectangles and squares piled on each other with each end poking out a bit further than the one just above, and we observe that the total area of this portrait is given by the relation

$$\begin{aligned}\mathcal{A}_1 &= r_1h_2 + (r_1 + r_2)h_3 + \cdots (r_1 + r_2 + \cdots + r_{n-2})h_{n-1} + (r_1 + r_2 + \cdots + r_{n-1})h_n \\ &= r_1(h_2 + h_3 + \cdots h_n) + r_2(h_3 + h_4 + \cdots + h_n) + \cdots + r_{n-2}(h_{n-1} + h_n) + r_{n-1}h_n \\ &= \sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}. \end{aligned} \tag{2.1}$$

On the other hand, we observe that the area of this portrait is the same as the difference of the area of triangle ΔABC and the sum of the areas of triangles $\Delta A_jB_jC_j$ for $j = 1, 2, \ldots, n$. That is

$$\mathcal{A}_1 = \frac{1}{2}rh - \frac{1}{2}\sum_{j=1}^{n} r_jh_j. \tag{2.2}$$

This completes the first part of the argument. For the second part, along the hypotenuse, let us construct small pieces of triangle, each of base and height (r_i, h_i) $(i = 1, 2 \ldots, n)$ so that the trapezoid and the one triangle formed by partitioning becomes rectangles and squares. We observe also that this construction satisfies the relation

$$(r^2 + h^2)^{1/2} = \sum_{i=1}^{n} (r_i^2 + h_i^2)^{1/2},$$

Now, we compute the area of the triangle in two different ways. By direct strategy, we have that the area of the triangle, denoted $\mathcal{A}$, is given by

$$\mathcal{A} = 1/2\Bigg(\sum_{i=1}^{n} r_i\Bigg)\Bigg(\sum_{i=1}^{n} h_i\Bigg). \tag{2.3}$$

On the other hand, we compute the area of the triangle by computing the area of each trapezium and the one remaining triangle and sum them together. That is,

$$\mathcal{A} = h_n/2\Bigg(\sum_{i=1}^{n} r_i + \sum_{i=1}^{n-1} r_i\Bigg) + h_{n-1}/2\Bigg(\sum_{i=1}^{n-1} r_i + \sum_{i=1}^{n-2} r_i\Bigg) + \cdots + 1/2r_1h_1. \tag{2.4}$$

By comparing equation (2.1) with equation (2.2), and comparing equation (2.3) with equation (2.4) in the resulting equation the result follows immediately. □

Corollary 2.2. *Let $f : \mathbb{N} \longrightarrow \mathbb{C}$, then we have the decomposition*

$$\sum_{n \leq x-1} \sum_{j \leq x-n} f(n)f(n+j) = \sum_{2 \leq n \leq x} f(n) \sum_{m \leq n-1} f(m).$$

Proof. Let us take $f(j) = r_j = h_j$ in Theorem 2.1, then we denote by $\mathcal{G}$ the partial sums

$$\mathcal{G} = \sum_{j=1}^{n} f(j)$$

and we notice that

$$\begin{aligned}\sum_{j=1}^{n}\sqrt{(h_j^2+r_j^2)} &= \sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}\\ &= \sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}\\ &= \sqrt{2}\sum_{j=1}^{n}f(j).\end{aligned}$$

Since $\sqrt{(\mathcal{G}^2+\mathcal{G}^2)}=\mathcal{G}\sqrt{2}=\sqrt{2}\sum\limits_{j=1}^{n}f(j)$ our choice of sequence is valid and, therefore the decomposition is valid for any arithmetic function. □

Theorem 2.3. *Let $f:\mathbb{N}\longrightarrow\mathbb{R}^{+}$, a real-valued function. If*

$$\sum_{n\leq x}f(n)f(n+l_0)>0$$

then there exist some constant $\mathcal{C}:=\mathcal{C}(l_0)>0$ fixed such that

$$\sum_{n\leq x}f(n)f(n+l_0)\geq\frac{1}{\mathcal{C}(l_0)x}\sum_{2\leq n\leq x}f(n)\sum_{m\leq n-1}f(m).$$

Proof. By Theorem 2.1, we obtain the identity by taking $f(j)=r_j=h_j$

$$\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j)=\sum_{2\leq n\leq x}f(n)\sum_{m\leq n-1}f(m).$$

It follows that

$$\begin{aligned}\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j) &\leq \sum_{n\leq x-1}\sum_{j<x}f(n)f(n+j)\\ &= \sum_{n\leq x}f(n)f(n+1)+\sum_{n\leq x}f(n)f(n+2)\\ &+\cdots\sum_{n\leq x}f(n)f(n+l_0)+\cdots\sum_{n\leq x}f(n)f(n+x)\\ &\leq |\mathcal{M}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)\\ &+|\mathcal{N}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)\\ &+\cdots+\sum_{n\leq x}f(n)f(n+l_0)+\cdots+|\mathcal{R}(l_0)|\sum_{n\leq x}f(n)f(n+l_0)\\ &= \bigg(|\mathcal{M}(l_0)|+|\mathcal{N}(l_0)|+\cdots+1\\ &+\cdots+|\mathcal{R}(l_0)|\bigg)\sum_{n\leq x}f(n)f(n+l_0)\\ &\leq \mathcal{C}(l_0)x\sum_{n<x}f(n)f(n+l_0)\end{aligned}$$

where $\max\{|\mathcal{M}(l_0)|, |\mathcal{N}(l_0)|, \ldots, |\mathcal{R}(l_0)|\} = \mathcal{C}(l_0)$. By inverting this inequality, the result follows immediately. □

The nature of the implicit constant $\mathcal{C}(l_0)$ could also depend on the structure of the function we are being given. The von mangoldt function, contrary to many class of arithmetic functions, has a relatively small such constant. This behaviour stems from the fact that the Von-mangoldt function is defined on the prime powers. Thus one would expect most terms of sums of the form

$$\sum_{n\leq x-1}\sum_{j\leq x-n}\Lambda(n)\Lambda(n+j)$$

to fall off when j is odd for any prime power $n = p^k$ such that $j + p^k \neq 2^s$.

3. Main result

We are now ready to prove the main result of this paper. We assemble the tools we have developed thus far to solve the problem.

Theorem 3.1. *Let $\mathbb{P}$ denotes the set of all prime numbers, then we have the estimate*

$$\#\{p \leq x \mid p+4, p \in \mathbb{P}\setminus\{2\}\} \geq (1+o(1))\frac{1}{2\mathcal{D}(4)}\frac{x}{\log^2 x}$$

where $\mathcal{D}(4) > 0$ fixed.

Proof. Let us consider the function $\vartheta : \mathbb{N} \longrightarrow \mathbb{R}^+$ defined as

$$\vartheta(n) := \begin{cases} \log p & \textbf{if} \quad n = p \in \mathbb{P} \\ 0 & \textbf{otherwise} \end{cases}$$

so that by virtue of Corollary 2.2 we obtain the lower bound

$$\sum_{n\leq x}\vartheta(n)\vartheta(n+4) \geq \frac{1}{x\mathcal{D}}\sum_{2\leq n\leq x}\vartheta(n)\sum_{m\leq n-1}\vartheta(m) \tag{3.1}$$

for $\mathcal{D} := \mathcal{D}(4) > 0$ fixed. Now using the weaker estimate found in the literature [2]

$$\sum_{n\leq x}\vartheta(n) = (1+o(1))x$$

we obtain the following estimates by an appeal to summation by parts

$$\begin{aligned}\sum_{2\leq n\leq x}\vartheta(n)\sum_{m\leq n-1}\vartheta(m) &= (1+o(1))\sum_{2\leq n\leq x}\vartheta(n)n\\ &= (1+o(1))x\sum_{2\leq n\leq x}\theta(n)-(1+o(1))\int_{2}^{x}\Bigg(\sum_{2\leq n\leq t}\vartheta(n)\Bigg)dt\\ &= (1+o(1))x^2-(1+o(1))\int_{2}^{x}(1+o(1))tdt\\ &= (1+o(1))x^2-(1+o(1))\frac{x^2}{2}+O(1)\\ &= (1+o(1))\frac{x^2}{2}. \end{aligned} \tag{3.2}$$

By plugging (3.2) into (3.1) we obtain the estimate

$$\begin{aligned}\sum_{n\leq x}\vartheta(n)\vartheta(n+4) &\geq \frac{1}{x\mathcal{D}}(1+o(1))\frac{x^2}{2}\\ &= (1+o(1))\frac{1}{2\mathcal{D}}x.\end{aligned}$$

On the other hand, we can write

$$\begin{aligned}\sum_{n\leq x}\vartheta(n)\vartheta(n+4) &= \sum_{\substack{p\leq x\\ p+4,p\in\mathbb{P}\setminus\{2\}}}\log p\log(p+4)\\ &\approx \sum_{\substack{p\leq x\\ p+4,p\in\mathbb{P}\setminus\{2\}}}\log^2 p\end{aligned}$$

so that by an application of partial summation we have

$$\sum_{\substack{p\leq x\\ p+4,p\in\mathbb{P}\setminus\{2\}}}\log^2 p\leq \log^2 x\sum_{\substack{p\leq x\\ p+4,p\in\mathbb{P}\setminus\{2\}}}1. \tag{3.3}$$

By combining (3.2), (3.1) and (3.3) the lower bound follows as a consequence. □

Corollary 3.2. *There are infinitely many primes $p\in\mathbb{P}\setminus\{2\}$ such that $p+4\in\mathbb{P}$.*

Proof. Appealing to Theorem 3.1, we have the lower bound

$$\#\{p\leq x \mid p+4,p\in\mathbb{P}\setminus\{2\}\}\geq (1+o(1))\frac{1}{2\mathcal{D}(4)}\frac{x}{\log^2 x}$$

where $\mathcal{D}(4)>0$ fixed. By taking limits $x\longrightarrow\infty$ on both sides, we have

$$\lim_{x\longrightarrow\infty}\#\{p\leq x \mid p+4,p\in\mathbb{P}\setminus\{2\}\}=\infty$$

thereby ending the proof. □

Remark 3.3. It is worth noting that with Theorem 3.1 lower bound, we have proven the infinity of cousin primes. This method is useful in terms of generality because it may be used to find lower bounds for a wide range of correlated sums of the type.

$$\sum_{n\leq x} G(n)G(n+k)$$

for a uniform $1 \leq k \leq x$.

4. **Conclusion**

The method used to prove the twin prime conjecture in this study is straightforward and elegant. This method can also be used to develop an estimate for universal sums of the form in the spirit of addressing the binary Goldbach conjecture

$$\sum_{n\leq x} G(n)G(x-n)$$

which we do not pursue in this paper.

References

1. T. Agama, *The area method and applications*, arXiv preprint arXiv:1903.09257, 2019.
2. Montgomery, Hugh L *Topics in multiplicative number theory*, vol. 227, Springer, 2006.
3. T. Agama, *A proof of the twin prime conjecture*, arXiv preprint arXiv:1707.03265, 2017.

Department of Mathematics, African Institute for mathematical sciences, Ghana.
E-mail address: `Theophilus@aims.edu.gh/emperordagama@yahoo.com`