A PROOF OF THE TWIN PRIME CONJECTURE

T. AGAMA

Abstract. In this paper we prove the twin prime conjecture by showing that

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

where $\mathcal{C}:=\mathcal{C}(2)>0$ fixed and $\mathbb{P}$ is the set of all prime numbers. In particular it follows that

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

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

Let $G:\mathbb{N}\longrightarrow \mathbb{C}$ and consider correlated sums of the forms

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

and

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

where $1\leq l\leq x$. It is generally not easy to control sums of these forms, and unfortunately many of the open problems in number theory can be phrased in this manner. The twin prime conjecture, conjectured by De polignac, is one of the important open problems in number theory and the whole of mathematics. It can be expressed in the form

$$\sum_{n\leq x}\vartheta(n)\vartheta(n+2)$$

where

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

Date: August 17, 2021.
2010 *Mathematics Subject Classification.* Primary 11Pxx, 11Bxx; Secondary 11Axx, 11Gxx.
Key words and phrases. prime; twin; area method.

so that by exploiting summation by parts we can recover an estimate for the sum

$$\sum_{\substack{p\leq x\\ p+2\in\mathbb{P}}} 1$$

where $\mathbb{P}$ denotes the set of all prime numbers, and it is the case that obtaining a non-trivial lower bound for this correlation solves the twin prime conjecture. There are a good number of techniques in the literature for studying such sums, like the circle method of Hardy and little-wood, the sieve method and many others.
In this paper, we introduce the area method. This method can also be used to control correlated sums of the form above. The novelty of this method is that it allows us to write any of these correlated sums as a double sum, which is much easier to estimate using existing tools such as the summation by part formula. As an application we obtain the result:

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

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

where $\mathcal{D}(2)>0$ 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

In this section we introduce and develop a fundamental method for solving problems related to correlations of arithmetic functions. This method is fundamental in the sense that it uses the properties of four main geometric shapes, namely the triangle, the trapezium, the rectangle and the square. The basic identity we will derive is an outgrowth of exploiting the areas of these shapes and putting them together in a unified manner.

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

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

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 $\sqrt{(\mathcal{G}^2 + \mathcal{G}^2)} = \mathcal{G}\sqrt{2} = \sqrt{2} \sum_{j=1}^{n} f(j)$ 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 twin prime conjecture. 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*

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

where $\mathcal{D}(2) > 0$ 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

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

for $\mathcal{D} := \mathcal{D}(2) > 0$ fixed. Now using the weaker estimate found in the literature [1]

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

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

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

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

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

On the other hand, we can write

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

so that by an application of partial summation we have

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

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

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

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

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

where $\mathcal{D}(2) > 0$ fixed. By taking limits $x \longrightarrow \infty$ on both sides, we have

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

thereby ending the proof. □

Remark 3.3. It is important to remark that with the lower bound in Theorem 3.1, we have solved the twin prime conjecture. This method not only does it solve the twin prime conjecture, but is good in terms of its generality, for it can be used to obtain lower bounds for a general class of correlated sums of the form

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

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

4. **Conclusion**

The method adopted in this paper to prove the twin prime conjecture is simple and very elegant. In the spirit of solving the binary Goldbach conjecture, this method can also be exploited to develop an estimate for general sums of the form

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

which we do not pursue in this paper.

References

1. Montgomery, Hugh L *Topics in multiplicative number theory*, vol. 227, Springer, 2006.

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