Preprints.org

Article

On the Additive and Subtractive Representation of Even Numbers from Primes

Ali Shehu * and Jetmira Uka

Article

On the Additive and Subtractive Representation of Even Numbers from Primes

Ali Shehu [1],* and Jetmira Uka [2]

[1]   Alma Mater: Queen Mary University of London, United Kingdom
[2]   Brunel University London, United Kingdom jetmira.uka@brunel.ac.uk
*   Correspondence: ali.shehu@hotmail.co.uk

Abstract: We demonstrate a new quantitative method to the sieve of Eratosthenes, which is an alternative to the sieve of Legendre. In this method, every element of a given set is sifted out once only, and therefore, this method is free of the Möbius function and of the parity barrier. Using this method, we prove that every sufficiently large even number is the sum of two primes, and that every even number is the difference of two primes in infinitely many ways.

Keywords: Sieve of Eratosthenes; Goldbach conjecture; Polignac conjecture; Twin Prime conjecture

1. Introduction

1.1. Representing Even Numbers from Primes

Since the set of prime numbers is infinite, and since all prime numbers greater than 2 are odd numbers, then one knows immediately that there are infinitely many even numbers that can be represented as the sum of two primes, and, infinitely many even numbers that can be represented as the difference of two primes. Having said that, the following questions then naturally arise:

(i)   Can every even number be represented as the sum of two primes?
(ii)   Can every even number be represented as the difference of two primes?
(iii)   Can any even number, or indeed all even numbers, be represented as the difference of two primes, in infinitely many ways?
(iv)   If questions $'(ii)'$ and $'(iii)'$ are answered in the affirmative, can they also hold for consecutive primes, in representing even numbers greater than 4 as the difference of two primes?

The earliest known record to have posed question $'(i)'$, known as the Goldbach conjecture, dates back to 1742, in a correspondence between C. Goldbach and L. Euler, where it is propositioned that every even number greater than 2 can be represented as the sum of two primes [1]. The Goldbach conjecture has more lately become known as the binary Goldbach conjecture, or the strong Goldbach conjecture, in order to distinguish it from the ternary Goldbach conjecture, or the weak Goldbach conjecture, which states that every odd number greater than 5 can be represented as the sum of three primes. The binary Goldbach conjecture has to date been shown empirically to hold for every even number up to $4 \cdot 10^{18}$ [2], however, a rigorous proof, or disproof, remains elusive.

Nevertheless, some related theoretical results to the binary Goldbach conjecture have been achieved, the closest of which is due to J. R. Chen, who in 1973, proved that every sufficiently large even number can be represented as the sum of a prime and another prime or a semiprime, that is, the product of two primes [3] [4] [5] [6]. On the other hand, significant results have been achieved for the ternary Goldbach conjecture, culminating with a proof given by H. Helfgott in 2014 [7] [8] [9].

The earliest known record to have posed questions $'(ii) - (iv)'$ are due to Alphonse de Polignac, who in 1849, conjectured that every even number can be represented as the difference of two consecutive primes, in infinitely many ways [10]. The most significant special case of Polignac's conjecture, is the so-called 'twin prime conjecture', which comprises of the number 2 being represented as the difference of two primes, in infinitely many ways. The twin prime conjecture is hugely supported

by empirical data, where, over the past few decades, increasingly large twin prime pairs have been found to exist [11], with the current world record for a twin prime pair, found in the year 2016, standing at 388, 342 decimal digits long [12].

As with the binary Goldbach Conjecture, the closest theoretical result to Polignac's conjecture is given by J. R. Chen, who in the same article, proved that every even number can be represented as the difference of a prime and another prime or semiprime, that is, the product of at most two primes [6]. More recently, D. A. Goldston, J. Pintz, and C. Y. Yildirim, introduced the now known 'GPY method', which uses approximations to the prime k-tuples conjecture, to study small numbers that can be represented as the difference of two primes [13]. In 2013, Yitang Zhang built on the GPY method, thereby proving for the first time the existence of some even number less than $7 \cdot 10^7$, which can be represented as the difference of two primes in infinitely many ways [14]. Within a year after Zhang's result, J. Maynard presented an independent proof that lowered the bound to 600, which, by assuming the Elliott-Halberstam conjecture, could be further lowered down to 12 [15]. With some refinements to Zhang's method and a combination of that with Maynard's approach, the bound was lowered to 246 unconditionally, by an on-line collaborative project known as *Polymath 8*, organised by T. Tao [16].

1.2. Sieve Theory

Sieve theory is a technique for distinguishing specific subsets of integers, amongst the set of natural numbers. Sieve theory began with Eratosthenes of Cyrene (276-194 B.C.), who constructed a method with which one could isolate the subset of prime numbers, from the set of natural numbers [17]. It starts by first crossing the multiples of 2 in the number line up to x, then the multiples of 3, then the multiples of 5, and then the multiples of all the primes up to $\sqrt{x}$. If an integer $n > 1$ is not divisible by any prime $p \leq \sqrt{x}$, then n is necessarily a prime. Upon completion of the sieve, one has

$$\# \{p \in \mathbb{P} : p \leq x\} = \# \left\{n \leq x : p \nmid n, p \leq \sqrt{x}\right\}. \tag{1}$$

Having the means of constructing the complete subset of prime numbers, from the set of natural numbers up to x, one would be naturally interested in quantifying these primes. To this end, there are fundamentally two approaches that one could take, in order to quantify the set of prime numbers generated by the sieve of Eratosthenes. We describe each approach, as follows.

(I) In the first approach, one quantifies the proportion of integers that are sifted out at each round of the sieve, as

$$\frac{\#\{\text{multiples of } p \text{ up to } x\}}{\#\{\text{integers up to } x\}}.$$

The appeal of this approach lies in the straightforward quantitative definition of both the subset of the multiples of p up to x and the complete set of integers up to x. However, the challenging aspect of this approach arises from the fact that integers which have multiple prime factors are necessarily counted at multiple rounds of the sieve, which must be accounted for. This is the approach taken by A. Legendre (1752-1833), who was the first to turn the sieve of Eratosthenes into a quantitative technique, and this has been the approach taken since.

In the sieve of Legendre, one counts the integers that are sifted out at each round, and then one subtracts those that are counted at multiple rounds, according to the multiplicity of times that this has happened, as follows

$$[x] - \sum_{p \leq \sqrt{x}} \left[\frac{x}{p}\right] + \sum_{p_1 < p_2 \leq \sqrt{x}} \left[\frac{x}{p_1 p_2}\right] - \sum_{p_1 < p_2 < p_3 \leq \sqrt{x}} \left[\frac{x}{p_1 p_2 p_3}\right] + - \cdots ; \tag{2}$$

from which, one obtains Legendre's formula

$$\pi\left(x\right) - \pi\left(\sqrt{x}\right) + 1 = \sum_{\substack{d \\ p|d \rightarrow p \leq \sqrt{x}}} \mu\left(d\right)\left[\frac{x}{d}\right], \tag{3}$$

where $\mu\left(d\right)$ is the Möbius function, introduced by A. F. Möbius (1790–1868), and defined as

$$\mu\left(n\right) := \begin{cases} 1 & \text{if } x \text{ is square-free and has an even number of prime factors,} \\ -1 & \text{if } x \text{ is square-free and has an odd number of prime factors,} \\ 0 & \text{if } x \text{ is not square-free.} \end{cases} \tag{4}$$

In efforts to evaluate the right-hand side of 3, one has

$$\pi\left(x\right) - \pi\left(\sqrt{x}\right) + 1 = x\sum_{d}\frac{\mu\left(d\right)}{d} + R = x\prod_{p=p_1}^{p \leq \sqrt{x}}\frac{p-1}{p} + R, \tag{5}$$

where the remainder R is

$$R = -\sum_{d}\mu\left(d\right)\left\{\frac{x}{d}\right\}, \tag{6}$$

which doubles at each round of the sieve, and thus quickly becomes larger than x.

In modern sieve methods, one tries to replace $\mu\left(n\right)$ by a function $\Lambda = \left(\lambda_d\right)$, referred to as the "sieve weights", which mimics the $\mu\left(n\right)$ and gives satisfactory estimates on upper bounds, lower bounds, and asymptotics for smoother number sets such as almost primes, which leads to upper bounds for primes. However, obtaining lower bounds for primes has proved much more difficult, thus leaving Goldbach's and Polignac's conjectures out of reach. This has been due to the so-called "parity barrier", where the sieve struggles to distinguish integers with an odd number of prime factors from integers with an even number of prime factors. Although the parity barrier has been broken for certain specific sequences, it still remains the fundamental obstacle in the treatment of primes via this approach of sieve theory.

(II) In the second approach, one quantifies the proportion of integers that are sifted out at each round of the sieve, as

$$\frac{\#\{\text{multiples of } p \text{ up to } x \text{ that survived the preceding rounds}\}}{\#\{\text{integers up to } x \text{ that survived the preceding rounds}\}}.$$

This means that every integer is sifted out according to its least prime factor, out of a set of integers whose least prime factors are greater than, or equal to, the least prime factors of the integers being sifted out at the given round, with the exception of the number 1. This is best illustrated by the following example.

In the 1st round of the sieve, one quantifies the subset of the multiples of 2 up to 26, as a ratio over the set of all integers up to 26. That is,

$$\frac{\#\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26\}}{\#\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26\}}.$$

In the 2nd round, one quantifies the subset of the multiples of 3 that survived the preceding round of the sieve, as a ratio over the set of all the integers that survived the preceding round of the sieve. That is,

$$\frac{\#\{3, 9, 15, 21\}}{\#\{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25\}}.$$

In the 3rd round, one quantifies the subset of the multiples of 5 that survived the preceding rounds of the sieve, as a ratio over the set of all the integers that survived the preceding rounds of the sieve. That is,

$$\frac{\#\{5, 25\}}{\#\{1, 5, 7, 11, 13, 17, 19, 23, 25\}}.$$

Since that completes the sieve, then the subset of integers that survives the sieve is

$$\{1, 7, 11, 13, 17, 19, 23\},$$

which consists of the number 1, and the complete set of primes in the interval $[\sqrt{26}, 26\,]$.

The appeal of this approach lies in the fact that every integer is sifted out only once; and therefore, one does not have the issue of certain integers appearing at multiple rounds of the sieve, which one has with the first approach, as described above. However, the challenging aspect of this approach is the difficulty that one has in defining quantitatively the two subsets of integers that survive the sieve up to the given round. Nevertheless, since at each round of the sieve, the subset of the multiples of p that survive the preceding rounds consists of the complete set of integers up to x for which p is the least prime factor; and, the subset of all the integers that survive the preceding rounds contains the complete set of primes in the interval $[\sqrt{x}, x\,]$; then, one is able to define quantitatively the two sets of integers, at least to the extent where one can determine upper and/or lower bounds.

1.3. Our Results

In the present paper, we take the second approach as outlined above, in order to describe quantitatively the sieve of Eratosthenes. On the condition that x is sufficiently large; at each round of the sieving of integers, we define quantitatively both the subset of the multiples of p up to x that survive the preceding rounds of the sieve and the subset of all the integers up to x that survive the preceding rounds of the sieve; to the extent where we are then able to determine an upper bound to the proportion of integers that are sifted out at each round of the sieve. To the best of our knowledge, we are the first to take this approach for these purposes.

Definition 1 (Additive representation)**.** *Let $a, b \in \mathbb{N}$, let $a \le b$, and let x be an even number, then we say that $a + b$ is an additive representation of x, if $a + b = x$.*

Definition 2 (Subtractive representation)**.** *Let $a, b \in \mathbb{N}$, let $a < b$, and let k be an even number, then we say that $b - a$ is a subtractive representation of k, if $b - a = k$.*

We then extend our approach taken in the sieving of integers, to the sieving of representations, whereby we quantify the subset of the additive representations of x, and the subset of the subtractive representations of k up to x, that survive the sieve of Eratosthenes. As with the sieving of integers, we quantify the proportion of representations that are sifted out at each round of the sieve, as

$$\frac{\#\{\text{representations, containing multiples of } p \text{ up to } x, \text{ that survived the preceding rounds}\}}{\#\{\text{representations, containing integers up to } x, \text{ that survived the preceding rounds}\}}.$$

This means that additive representations, and subtractive representations, are sifted out according to either a or b, depending on whose least prime factor is the smaller (or equal to) of the two. This allows for every representation that contains at least one multiple of p to be sifted out once only, while every representation that does not contain multiples of p, survives the sieve. Therefore, upon completion of the sieve, we have a subset of representations where $a = 1$ or a prime in the interval $[\sqrt{x}, x\,]$, and b is a prime in the interval $[\sqrt{x}, x\,]$; with the quantity of representations where $a = 1$ being at most one.

By relating the sieving of representations to the sieving of integers for the same x, we determine an upper bound to the proportion of additive representations that are sifted out at each round of the sieve, and an upper bound to the proportion of subtractive representations that are sifted out at each round of the sieve, on the condition that x is sufficiently large. Following on from the upper bounds, we then determine a positive lower bound to the subset of additive representations that survive the sieve, where both a and b are primes in the interval $[\sqrt{x}, x]$, and a positive lower bound to the subset of subtractive representations that survive the sieve, where both a and b are primes in the interval $[\sqrt{x}, x]$. Therefore, we prove the following:

Theorem 1. *Every sufficiently large even number is the sum of two primes.*

Theorem 2. *Every even number is the difference of two primes in infinitely many ways.*

Theorem 1 partially addresses question $'(i)'$ as posed above; and, to date is the closest theoretical result to the binary Goldbach conjecture. Theorem 2 fully addresses questions $'(ii) - (iii)'$ as posed above; fully establishes the Twin Prime conjecture; and, to date is the closest theoretical result to the Polignac conjecture.

2. Notation

$\mathbb{N}$ and $\mathbb{P}$: the set of natural numbers and the set of prime numbers respectively.

x and k: even numbers. p and q: prime numbers. p_n: the nth prime number. We have $p_m < p_n$.

w: a sufficiently large positive integer, not necessarily the same at every occurrence.

r_z: the zth integer in the sequence $r := \{a \in \mathbb{N} : 2 \nmid a, 3 \nmid a\}$.

$\alpha_1, \alpha_2, \alpha_m, \alpha_n, \eta_n, h_n, l_n$: positive real numbers.

$\beta_1, \beta_2, \beta_m, \beta_n, \theta_n, \gamma_1, \gamma_2, \gamma_m, \gamma_n, \nu_n$: non-negative real numbers.

$\pi(a) := |\{p : p \leq a\}|$.

$\pi[a,b] := |\{p : a \leq p \leq b\}|$.

$g(x) := |\{p : (x - p) \in \mathbb{P}, \sqrt{x} \leq p \leq (x - p) \leq x\}|$.

$\pi_k(x) := |\{p : (p + k) \in \mathbb{P}, \sqrt{x} \leq p \leq (p + k) \leq x\}|$.

$\pi_k(x)' := |\{p : (p + k) \in \mathbb{P}, 2 \leq p \leq (p + k) \leq x\}|$.

$\mathcal{A} := \{a \in \mathbb{N} : a \leq x\}$.

$\mathcal{A}^p := \{a \in \mathcal{A} : q \nmid a, q < p\}$.

$\mathcal{A}_p := \{a \in \mathcal{A} : p \mid a\}$.

$\mathcal{A}_{qp} := \{a \in \mathcal{A} : q \mid a \text{ and } p \mid a\}$.

$S(\mathcal{A}, p) := \{a \in \mathcal{A} : p \nmid a\}$.

$\mathcal{B} := \{b \in \mathbb{N} : b \leq (x - b)\}$.

$\mathcal{B}^p := \{b \in \mathcal{B} : q \nmid b \text{ and } q \nmid (x - b), q < p\}$.

$\mathcal{B}_p := \{b \in \mathcal{B} : p \mid b \text{ or } p \mid (x - b)\}$.

$\mathcal{B}_{qp} := \left\{b \in \mathcal{B} : \left[q \mid b \text{ or } q \mid (x - b)\right] \text{ and } \left[p \mid b \text{ or } p \mid (x - b)\right]\right\}$.

$S(\mathcal{B}, p) := \{b \in \mathcal{B} : p \nmid b \text{ and } p \nmid (x - b)\}$.

$\mathcal{C} := \left\{c \in \mathbb{N} : (c + k) \leq x, k \leq \frac{x}{2}\right\}$.

$\mathcal{C}^p := \{c \in \mathcal{C} : q \nmid c \text{ and } q \nmid (c + k), q < p\}$.

$\mathcal{C}_p := \{c \in \mathcal{C} : p \mid c \text{ or } p \mid (c + k)\}$.

$\mathcal{C}_{qp} := \left\{c \in \mathcal{C} : \left[q \mid c \text{ or } q \mid (c + k)\right] \text{ and } \left[p \mid c \text{ or } p \mid (c + k)\right]\right\}$.

$S(\mathcal{C}, p) := \{c \in \mathcal{C} : p \nmid c \text{ and } p \nmid (c + k)\}$.

$\mathcal{D} := \left[d^{\wedge}e \in \mathbb{N} : d \leq x, k \leq \frac{x}{2}, e = \begin{cases} 2, & \text{if } (1 + k) \leq d \leq (x - k), \\ 1, & \text{otherwise,} \end{cases}\right]$; that is, $\mathcal{D}$ is a multiset, with e being the multiplicity of d.

$\mathcal{D}_p := [d \in \mathcal{D} : p \mid d]$. Any other notation used, is defined at the point of use.

3. Sieve Outline

3.1. Sieving Integers

Let

$$|\mathcal{A}_{p_n}^{p_n}| = \frac{\alpha_n}{p_n} \cdot |\mathcal{A}^{p_n}|, \tag{7}$$

then

$$|S(\mathcal{A}^{p_n}, p_n)| = |\mathcal{A}^{p_n}| - \frac{\alpha_n}{p_n} \cdot |\mathcal{A}^{p_n}| = |\mathcal{A}^{p_{n+1}}|. \tag{8}$$

Since $|\mathcal{A}| = x$, then

$$|\mathcal{A}^{p_{n+1}}| = x \cdot \frac{p_1 - \alpha_1}{p_1} \cdot \frac{p_2 - \alpha_2}{p_2} \cdot \ldots \cdot \frac{p_n - \alpha_n}{p_n}. \tag{9}$$

Let $p_n \leq \sqrt{x}$, then

$$|\mathcal{A}^{p_{n+1}}| = \pi(x) - \pi(\sqrt{x}) + 1, \tag{10}$$

and therefore

$$\pi(x) - \pi(\sqrt{x}) + 1 = x \cdot \prod_{p_n \leq \sqrt{x}} \frac{p_n - \alpha_n}{p_n}. \tag{11}$$

Empirical data for the sieving of $\mathcal{A}$, as shown in Figure 1, suggest that α_n oscillates about 1, before approaching 0 as $p_n \to \sqrt{x}$. We determine an upper bound to α_n, for $p_n \leq \sqrt{x}$ and $x \geq w$.

(a) $x = L \cdot 10^8, \ p_n \leq 1 \cdot 10^3$ (b) $x = L \cdot 10^8, \ p_n \leq \sqrt{x}$
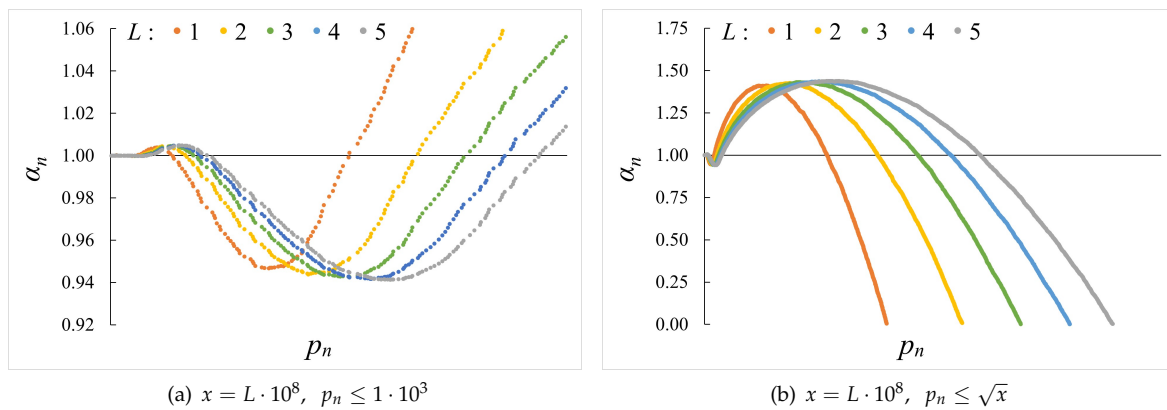
Figure 1. $\alpha_n = \frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \cdot p_n$, for various x.

3.2. Sieving Additive Representations

Let

$$|\mathcal{B}_{p_n}^{p_n}| = \frac{\beta_n}{p_n} \cdot |\mathcal{B}^{p_n}|, \tag{12}$$

then

$$|S(\mathcal{B}^{p_n}, p_n)| = |\mathcal{B}^{p_n}| - \frac{\beta_n}{p_n} \cdot |\mathcal{B}^{p_n}| = |\mathcal{B}^{p_{n+1}}|. \tag{13}$$

Since $|\mathcal{B}| = \frac{x}{2}$, then

$$|\mathcal{B}^{p_{n+1}}| = \frac{x}{2} \cdot \frac{p_1 - \beta_1}{p_1} \cdot \frac{p_2 - \beta_2}{p_2} \cdot \ldots \cdot \frac{p_n - \beta_n}{p_n}. \tag{14}$$

Let $p_n \leq \sqrt{x}$, then

$$|\mathcal{B}^{p_{n+1}}| = g(x) + u, \tag{15}$$

where $u \leq 1$; and therefore,

$$g(x) \geq \frac{x}{2} \cdot \prod_{p_n \leq \sqrt{x}} \frac{p_n - \beta_n}{p_n} - 1. \tag{16}$$

Empirical data for the sieving of $\mathcal{B}$ and for that of $\mathcal{A}$, as shown in Figure 2 for $p_n \leq \sqrt{x}$, suggest that

$$\frac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \approx \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \mid x, \\[3mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \nmid x, \end{cases} \tag{17}$$

which we generalise for $p_n \leq \sqrt{x}$. Through 17, we determine upper bounds to β_n for $p_n \leq \sqrt{x}$ and $x \geq w$. Following on from the upper bounds for β_n, we determine a positive lower bound to 16 for $x \geq w$, and thus we prove Theorem 1.

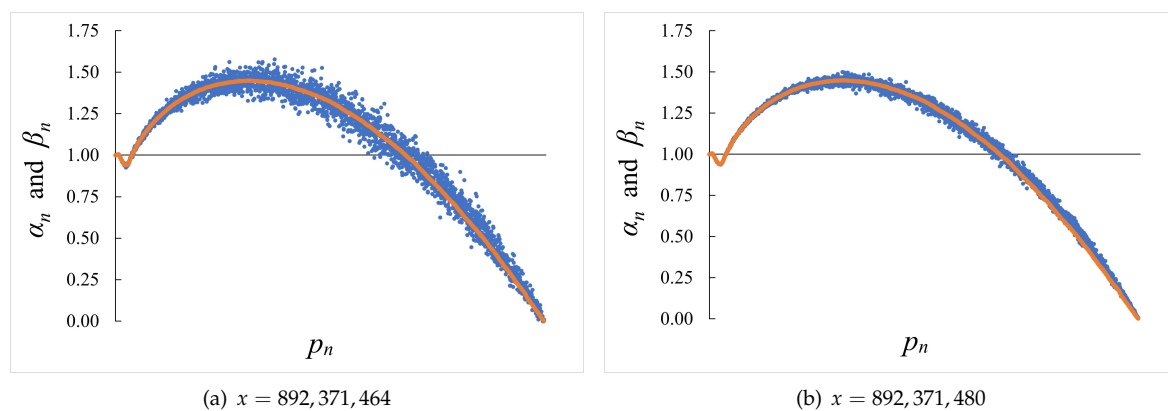(a) $x = 892,371,464$ (b) $x = 892,371,480$

Figure 2. $\alpha_n = \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \cdot p_n$ (red) and $\beta_n = \begin{cases} \dfrac{1}{1} \cdot \dfrac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \cdot p_n & \text{if } p_n \mid x, \\[3mm] \dfrac{1}{2} \cdot \dfrac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \cdot p_n & \text{if } p_n \nmid x, \end{cases}$ (blue).

3.3. Sieving subtractive representations

Let

$$|\mathcal{C}_{p_n}^{p_n}| = \frac{\gamma_n}{p_n} \cdot |\mathcal{C}^{p_n}|, \tag{18}$$

then

$$|S(\mathcal{C}^{p_n}, p_n)| = |\mathcal{C}^{p_n}| - \frac{\gamma_n}{p_n} \cdot |\mathcal{C}^{p_n}| = |\mathcal{C}^{p_{n+1}}|. \tag{19}$$

Since $|\mathcal{C}| = x - k$, then

$$|\mathcal{C}^{p_{n+1}}| = (x - k) \cdot \frac{p_1 - \gamma_1}{p_1} \cdot \frac{p_2 - \gamma_2}{p_2} \cdot \ldots \cdot \frac{p_n - \gamma_n}{p_n}. \tag{20}$$

Let $p_n \leq \sqrt{x}$, then

$$|\mathcal{C}^{p_{n+1}}| = \pi_k(x) + u, \tag{21}$$

where $u \leq 1$; and therefore,

$$\pi_k(x) \geq (x - k) \cdot \prod_{p_n \leq \sqrt{x}} \frac{p_n - \gamma_n}{p_n} - 1. \tag{22}$$

Empirical data for the sieving of $\mathcal{C}$ and for that of $\mathcal{A}$, as shown in Figure 3 for $p_n \leq \sqrt{x}$, suggest that

$$\frac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \approx \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \mid k, \\[4mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \nmid k, \end{cases} \tag{23}$$

which we generalise for $p_n \leq \sqrt{x}$. Through 23, we determine upper bounds to γ_n for $p_n \leq \sqrt{x}$ and $x \geq w$. Following on from the upper bounds for γ_n, we determine a positive lower bound to 22 for $x \geq w$, and thus we prove Theorem 2.

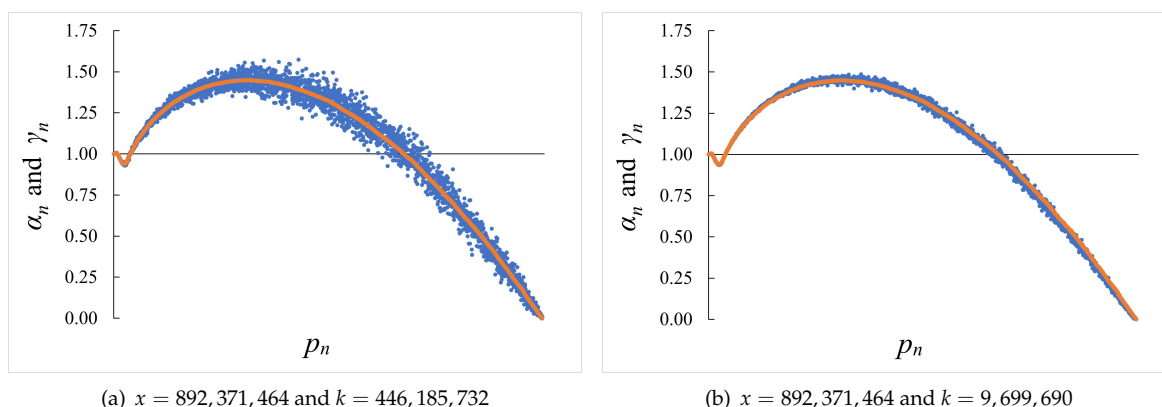(a) $x = 892,371,464$ and $k = 446,185,732$ (b) $x = 892,371,464$ and $k = 9,699,690$

Figure 3. $\alpha_n = \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \cdot p_n$ (red) and $\gamma_n = \begin{cases} \dfrac{1}{1} \cdot \dfrac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \cdot p_n & \text{if } p_n \mid k, \\[3mm] \dfrac{1}{2} \cdot \dfrac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \cdot p_n & \text{if } p_n \nmid k, \end{cases}$ (blue).

4. Preliminaries

Proposition 1. *Let $p_n \leq \sqrt{x}$. Since $\mathcal{A}_{p_n}^{p_n}$ consists of the complete set of integers ($\leq x$) for which p_n is the least prime factor, then due to the Fundamental Theorem of Arithmetic, we have*

$$|\mathcal{A}_{p_n}^{p_n}| \equiv \pi\left[p_n, p_n\right] + \pi\left[p_n, \frac{x}{p_n}\right] + \pi\left[p_n, \frac{x}{p_n \cdot p_n}\right] + \pi\left[p_{n+1}, \frac{x}{p_n \cdot p_{n+1}}\right] + \cdots, \tag{24}$$

where $\pi\left[p_n, \dfrac{x}{p_n \cdot p_n}\right] + \pi\left[p_{n+1}, \dfrac{x}{p_n \cdot p_{n+1}}\right] + \cdots = 0$, if $p_n > \sqrt[3]{x}$.

Proposition 2. *In the following, we describe the sieving process of $\mathcal{A}$ by p_m. Thus, let*

$$|\mathcal{A}_{p_m}^{p_m}| = \frac{\alpha_m}{p_m} \cdot |\mathcal{A}^{p_m}|, \tag{25}$$

and

$$|\mathcal{A}_{p_n}^{p_m}| = \frac{\eta_n}{p_n} \cdot |\mathcal{A}^{p_m}|. \tag{26}$$

Then, in light of distribution effects at finite range, we have

$$|\mathcal{A}_{p_m p_n}^{p_m}| \approx \frac{\alpha_m}{p_m} \cdot \frac{\eta_n}{p_n} \cdot |\mathcal{A}^{p_m}|, \tag{27}$$

and

$$|\mathcal{A}_{p_m p_n}^{p_m}| \approx \frac{\eta_n}{p_n} \cdot |\mathcal{A}_{p_m}^{p_m}| \approx \frac{\alpha_m}{p_m} \cdot |\mathcal{A}_{p_n}^{p_m}|. \tag{28}$$

Therefore, due to 25 we have

$$|S\left(\mathcal{A}^{p_m}, p_m\right)| = |\mathcal{A}^{p_m}| - \frac{\alpha_m}{p_m} \cdot |\mathcal{A}^{p_m}|, \tag{29}$$

and due to 28 we have

$$|S\left(\mathcal{A}_{p_n}^{p_m}, p_m\right)| \approx |\mathcal{A}_{p_n}^{p_m}| - \frac{\alpha_m}{p_m} \cdot |\mathcal{A}_{p_n}^{p_m}|. \tag{30}$$

Therefore, due to 26 we have

$$|\mathcal{A}_{p_n}^{p_{m+1}}| \approx \frac{\eta_n}{p_n} \cdot |\mathcal{A}^{p_{m+1}}|, \tag{31}$$

where

$$\lim_{|\mathcal{A}_{p_m p_n}^{p_m}| \to \infty} \frac{|\mathcal{A}_{p_n}^{p_{m+1}}|}{|\mathcal{A}^{p_{m+1}}|} = \frac{\eta_n}{p_n}. \tag{32}$$

Proposition 3. *Since*

$$\frac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|} \leq \frac{1}{p_n}; \tag{33}$$

and since $|\mathcal{A}_{p_m p_n}^{p_m}| \to \infty$ as $|\mathcal{A}_{p_n}^{p_n}| \to \infty$; then, in light of Proposition 2, we have

$$\lim_{|\mathcal{A}_{p_n}^{p_n}| \to \infty} \frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} = \frac{1}{p_n}. \tag{34}$$

Proposition 4. *Let $p_n \leq \sqrt{x}$, then we have*

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} = \frac{|\mathcal{A}_{p_n}^{p_n}|}{1 + \pi\left[\sqrt{x}, x\right] + |\mathcal{A}_{p_n}^{p_n}| + |\mathcal{E}|}, \tag{35}$$

where $\mathcal{E}$ is the set of the remaining multiples of the primes in the interval $\left[p_{n+1}, \sqrt{x}\right]$; and therefore, $|\mathcal{E}| \geq \pi\left[p_{n+1}, \sqrt{x}\right]$. Since x is an even number, then $\sqrt{x}$ is non-prime; and therefore, if $p_{n+1} < \sqrt{x}$, then

$$\pi\left[p_{n+1}, \sqrt{x}\right] + \pi\left[\sqrt{x}, x\right] = \pi\left[p_{n+1}, x\right]. \tag{36}$$

If $p_{n+1} > \sqrt{x}$, then $|\mathcal{E}| = 0$ and p_{n+1} is the 1st prime after the $\sqrt{x}$; where, the $\sqrt{x}$ is non-prime. Therefore; if $p_{n+1} > \sqrt{x}$, then $\pi\left[\sqrt{x}, x\right] = \pi\left[p_{n+1}, x\right]$. Therefore; if $p_n \leq \sqrt{x}$, then

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{1 + \pi\left[\sqrt{x}, x\right] + |\mathcal{A}_{p_n}^{p_n}| + |\mathcal{E}|} \leq \frac{|\mathcal{A}_{p_n}^{p_n}|}{1 + |\mathcal{A}_{p_n}^{p_n}| + \pi\left[p_{n+1}, x\right]} = \frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}_{p_n}^{p_n}| + \pi\left[p_n, x\right]}, \tag{37}$$

and therefore

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \leq \frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}_{p_n}^{p_n}| + \pi\left[p_n, x\right]}. \tag{38}$$

Proposition 5. *Since $p_n \leq \sqrt{x}$ and $|\mathcal{A}^{p_n}| \geq \pi\left[\sqrt{x}, x\right]$, then*

$$\frac{p_n}{|\mathcal{A}^{p_n}|} \leq \frac{\sqrt{x}}{\pi\left[\sqrt{x}, x\right]}; \tag{39}$$

where, due to the Prime Number Theorem, we have

$$\frac{\sqrt{x}}{\pi\left[\sqrt{x},x\right]} \sim \frac{x^{\frac{1}{2}}}{\dfrac{x}{\log x} - \dfrac{x^{\frac{1}{2}}}{\log x^{\frac{1}{2}}}} = \frac{\log x}{x^{\frac{1}{2}} - 2}. \tag{40}$$

Since

$$\lim_{x\to\infty} \frac{\log x}{x^{\frac{1}{2}} - 2} = 0, \tag{41}$$

then

$$\lim_{x\to\infty} \frac{\sqrt{x}}{\pi\left[\sqrt{x},x\right]} = 0, \tag{42}$$

and therefore

$$\lim_{x\to\infty} \frac{p_n}{|\mathcal{A}^{p_n}|} = 0. \tag{43}$$

Proposition 6. *Let $b \in \mathbb{N}$, then we have:* $\begin{cases} \text{if } p \mid x \text{ and } p \mid b, & \text{then } p \mid (x-b), \\ \text{if } p \nmid x \text{ and } p \mid b, & \text{then } p \nmid (x-b), \\ \text{if } p \nmid x \text{ and } p \mid (x-b), & \text{then } p \nmid b. \end{cases}$

Proposition 7. *Let $c \in \mathbb{N}$, then we have:* $\begin{cases} \text{if } p \mid k \text{ and } p \mid c, & \text{then } p \mid (c+k), \\ \text{if } p \nmid k \text{ and } p \mid c, & \text{then } p \nmid (c+k), \\ \text{if } p \nmid k \text{ and } p \mid (c+k), & \text{then } p \nmid c. \end{cases}$

Proposition 8. *In the following, we describe the sieving process of $\mathcal{B}$ by p_m; in relation to the sieving process of $\mathcal{A}$, also by p_m; as outlined in Proposition 2. Thus; let x, p_m, and p_n, be the same as in Proposition 2; and, in light of Proposition 6, let*

$$|\mathcal{B}_{p_m}^{p_m}| = \begin{cases} \dfrac{\beta_m}{p_m} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \mid x, \\[2ex] \dfrac{2\beta_m}{p_m} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \nmid x, \end{cases} \tag{44}$$

and

$$|\mathcal{B}_{p_n}^{p_m}| = \begin{cases} \dfrac{\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_n \mid x, \\[2ex] \dfrac{2\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_n \nmid x. \end{cases} \tag{45}$$

Then, in light of distribution effects at finite range, we have

$$|\mathcal{B}_{p_m p_n}^{p_m}| \approx \begin{cases} \dfrac{\beta_m}{p_m} \cdot \dfrac{\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \mid x \text{ and } p_n \mid x, \\[2ex] \dfrac{\beta_m}{p_m} \cdot \dfrac{2\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \mid x \text{ and } p_n \nmid x, \\[2ex] \dfrac{2\beta_m}{p_m} \cdot \dfrac{\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \nmid x \text{ and } p_n \mid x, \\[2ex] \dfrac{2\beta_m}{p_m} \cdot \dfrac{2\theta_n}{p_n} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \nmid x \text{ and } p_n \nmid x, \end{cases} \tag{46}$$

and

$$
|\mathcal{B}_{p_m p_n}^{p_m}| \approx
\begin{cases}
\dfrac{\theta_n}{p_n} \cdot |\mathcal{B}_{p_m}^{p_m}| \approx \dfrac{\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \mid x \text{ and } p_n \mid x, \\[2ex]
\dfrac{2\theta_n}{p_n} \cdot |\mathcal{B}_{p_m}^{p_m}| \approx \dfrac{\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \mid x \text{ and } p_n \nmid x, \\[2ex]
\dfrac{\theta_n}{p_n} \cdot |\mathcal{B}_{p_m}^{p_m}| \approx \dfrac{2\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \nmid x \text{ and } p_n \mid x, \\[2ex]
\dfrac{2\theta_n}{p_n} \cdot |\mathcal{B}_{p_m}^{p_m}| \approx \dfrac{2\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \nmid x \text{ and } p_n \nmid x.
\end{cases}
\tag{47}
$$

Therefore, due to 44 we have

$$
|S\left(\mathcal{B}^{p_m}, p_m\right)| =
\begin{cases}
|\mathcal{B}^{p_m}| - \dfrac{\beta_m}{p_m} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \mid x, \\[2ex]
|\mathcal{B}^{p_m}| - \dfrac{2\beta_m}{p_m} \cdot |\mathcal{B}^{p_m}|, & \text{if } p_m \nmid x,
\end{cases}
\tag{48}
$$

and due to 47 we have

$$
|S\left(\mathcal{B}_{p_n}^{p_m}, p_m\right)| \approx
\begin{cases}
|\mathcal{B}_{p_n}^{p_m}| - \dfrac{\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \mid x, \\[2ex]
|\mathcal{B}_{p_n}^{p_m}| - \dfrac{2\beta_m}{p_m} \cdot |\mathcal{B}_{p_n}^{p_m}|, & \text{if } p_m \nmid x.
\end{cases}
\tag{49}
$$

Therefore, if

$$
\frac{|\mathcal{B}_{p_n}^{p_m}|}{|\mathcal{B}^{p_m}|} = \frac{|\mathcal{A}_{p_n}^{p_m}|}{|\mathcal{A}^{p_m}|};
\tag{50}
$$

then, due to 48 and 49, in conjunction with 29 and 30, we have

$$
\frac{|\mathcal{B}_{p_n}^{p_{m+1}}|}{|\mathcal{B}^{p_{m+1}}|} \approx \frac{|\mathcal{A}_{p_n}^{p_{m+1}}|}{|\mathcal{A}^{p_{m+1}}|};
\tag{51}
$$

where,

$$
\lim_{|\mathcal{A}_{p_m p_n}^{p_m}| \to \infty, \, |\mathcal{B}_{p_m p_n}^{p_m}| \to \infty} \frac{|\mathcal{A}^{p_{m+1}}| \cdot |\mathcal{B}_{p_n}^{p_{m+1}}|}{|\mathcal{A}_{p_n}^{p_{m+1}}| \cdot |\mathcal{B}^{p_{m+1}}|} = 1.
\tag{52}
$$

Proposition 9. *In the following, we describe the sieving process of $\mathcal{C}$ by p_m; in relation to the sieving process of $\mathcal{A}$, also by p_m; as outlined in Proposition 2. Thus; let x, p_m, and p_n, be the same as in Proposition 2; and, in light of Proposition 7, let*

$$
|\mathcal{C}_{p_m}^{p_m}| =
\begin{cases}
\dfrac{\gamma_m}{p_m} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \mid k, \\[2ex]
\dfrac{2\gamma_m}{p_m} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \nmid k,
\end{cases}
\tag{53}
$$

and

$$
|\mathcal{C}_{p_n}^{p_m}| =
\begin{cases}
\dfrac{\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_n \mid k, \\[2ex]
\dfrac{2\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_n \nmid k.
\end{cases}
\tag{54}
$$

Then, in light of distribution effects at finite range, we have

$$|\mathcal{C}_{p_m p_n}^{p_m}| \approx \begin{cases} \dfrac{\gamma_m}{p_m} \cdot \dfrac{\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \mid k \text{ and } p_n \mid k, \\[3mm] \dfrac{\gamma_m}{p_m} \cdot \dfrac{2\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \mid k \text{ and } p_n \nmid k, \\[3mm] \dfrac{2\gamma_m}{p_m} \cdot \dfrac{\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \nmid k \text{ and } p_n \mid k, \\[3mm] \dfrac{2\gamma_m}{p_m} \cdot \dfrac{2\nu_n}{p_n} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \nmid k \text{ and } p_n \nmid k, \end{cases} \quad (55)$$

and

$$|\mathcal{C}_{p_m p_n}^{p_m}| \approx \begin{cases} \dfrac{\nu_n}{p_n} \cdot |\mathcal{C}_{p_m}^{p_m}| \approx \dfrac{\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \mid k \text{ and } p_n \mid k, \\[3mm] \dfrac{2\nu_n}{p_n} \cdot |\mathcal{C}_{p_m}^{p_m}| \approx \dfrac{\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \mid k \text{ and } p_n \nmid k, \\[3mm] \dfrac{\nu_n}{p_n} \cdot |\mathcal{C}_{p_m}^{p_m}| \approx \dfrac{2\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \nmid k \text{ and } p_n \mid k, \\[3mm] \dfrac{2\nu_n}{p_n} \cdot |\mathcal{C}_{p_m}^{p_m}| \approx \dfrac{2\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \nmid k \text{ and } p_n \nmid k. \end{cases} \quad (56)$$

Therefore, due to 53 we have

$$|S\left(\mathcal{C}^{p_m}, p_m\right)| = \begin{cases} |\mathcal{C}^{p_m}| - \dfrac{\gamma_m}{p_m} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \mid k, \\[3mm] |\mathcal{C}^{p_m}| - \dfrac{2\gamma_m}{p_m} \cdot |\mathcal{C}^{p_m}|, & \text{if } p_m \nmid k, \end{cases} \quad (57)$$

and due to 56 we have

$$|S\left(\mathcal{C}_{p_n}^{p_m}, p_m\right)| \approx \begin{cases} |\mathcal{C}_{p_n}^{p_m}| - \dfrac{\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \mid k, \\[3mm] |\mathcal{C}_{p_n}^{p_m}| - \dfrac{2\gamma_m}{p_m} \cdot |\mathcal{C}_{p_n}^{p_m}|, & \text{if } p_m \nmid k. \end{cases} \quad (58)$$

Therefore, if

$$\frac{|\mathcal{C}_{p_n}^{p_m}|}{|\mathcal{C}^{p_m}|} = \frac{|\mathcal{A}_{p_n}^{p_m}|}{|\mathcal{A}^{p_m}|}; \quad (59)$$

then, due to 57 and 58, in conjunction with 29 and 30, we have

$$\frac{|\mathcal{C}_{p_n}^{p_{m+1}}|}{|\mathcal{C}^{p_{m+1}}|} \approx \frac{|\mathcal{A}_{p_n}^{p_{m+1}}|}{|\mathcal{A}^{p_{m+1}}|}; \quad (60)$$

where,

$$\lim_{|\mathcal{A}_{p_m p_n}^{p_m}| \to \infty, \, |\mathcal{C}_{p_m p_n}^{p_m}| \to \infty} \frac{|\mathcal{A}^{p_{m+1}}| \cdot |\mathcal{C}_{p_n}^{p_{m+1}}|}{|\mathcal{A}_{p_n}^{p_{m+1}}| \cdot |\mathcal{C}^{p_{m+1}}|} = 1. \quad (61)$$

5. An Upper Bound on the Sifted Out Integers

In this section, we determine an upper bound to α_n, as follows.
Let $p_n \leq \sqrt[3]{x}$, then

$$\pi\left[p_n, \frac{x}{p_n}\right] \geq \pi\left(\frac{x}{\sqrt[3]{x}}\right) - \pi\left(\sqrt[3]{x}\right); \quad (62)$$

where, due to the Prime Number Theorem, we have

$$\pi\left(\frac{x}{\sqrt[3]{x}}\right) - \pi\left(\sqrt[3]{x}\right) \sim \frac{x^{\frac{2}{3}}}{\log x^{\frac{2}{3}}} - \frac{x^{\frac{1}{3}}}{\log x^{\frac{1}{3}}} = \frac{\frac{3}{2}x^{\frac{1}{3}}\left(x^{\frac{1}{3}} - 2\right)}{\log x}, \tag{63}$$

where

$$\frac{\frac{3}{2}x^{\frac{1}{3}}\left(x^{\frac{1}{3}} - 2\right)}{\log x} \geq \frac{x^{\frac{1}{3}}}{\log x}, \tag{64}$$

for $x \geq w$. Since $\dfrac{x^{\frac{1}{3}}}{\log x} \to \infty$ as $x \to \infty$, then $\pi\left[p_n, \dfrac{x}{p_n}\right] \to \infty$ as $x \to \infty$. Therefore, if $p_n \leq \sqrt[3]{x}$; then, from Proposition 1, we can see that $|\mathcal{A}_{p_n}^{p_n}| \to \infty$ as $x \to \infty$. Therefore, if $p_n \leq \sqrt[3]{x}$ and $x \geq w$; then, due to Proposition 3, we have

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \leq \frac{1}{p_n} \cdot 1.15. \tag{65}$$

It remains to determine an upper bound to α_n for $\sqrt[3]{x} < p_n \leq \sqrt{x}$, which we do as follows.

Let $\sqrt[3]{x} < p_n \leq \sqrt{x}$; then, due to Proposition 1, we have

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}_{p_n}^{p_n}| + \pi\left[p_n, x\right]} \equiv \frac{1 + \pi\left[p_n, \dfrac{x}{p_n}\right]}{1 + \pi\left[p_n, \dfrac{x}{p_n}\right] + \pi\left[p_n, x\right]} \leq \frac{1 + \pi\left[p_n, \dfrac{x}{p_n}\right]}{1 + \pi\left[p_n, x\right]}, \tag{66}$$

where

$$\frac{1 + \pi\left[p_n, \dfrac{x}{p_n}\right]}{1 + \pi\left[p_n, x\right]} \leq \frac{1 + \pi\left[p_n, \dfrac{x}{p_n}\right] + \pi\left(p_{n-2}\right)}{1 + \pi\left[p_n, x\right] + \pi\left(p_{n-2}\right)} = \frac{\pi\left(\dfrac{x}{p_n}\right)}{\pi\left(x\right)}, \tag{67}$$

for $x \geq w$. Therefore, if $\sqrt[3]{x} < p_n \leq \sqrt{x}$ and $x \geq w$; then, due to Proposition 4, we have

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \leq \frac{\pi\left(\dfrac{x}{p_n}\right)}{\pi\left(x\right)}, \tag{68}$$

which we quantify as follows.

Let $f(a) := \dfrac{\pi(a)}{a}$; then

$$\frac{\pi\left(\dfrac{x}{p_n}\right)}{\pi\left(x\right)} = \frac{1}{p_n} \cdot \frac{f\left(\dfrac{x}{p_n}\right)}{f(x)}. \tag{69}$$

Since $p_n \leq \sqrt{x}$, then

$$\frac{f\left(\dfrac{x}{p_n}\right)}{f(x)} \leq \frac{f\left(\sqrt{x}\right)}{f(x)}. \tag{70}$$

Due to the Prime Number Theorem, we have

$$\lim_{x \to \infty} \frac{f\left(\sqrt{x}\right)}{f(x)} = 2.00; \tag{71}$$

and therefore; if $x \geq w$, then

$$\frac{f\left(\sqrt{x}\right)}{f(x)} \leq 2.15. \tag{72}$$

Therefore; if $\sqrt[3]{x} < p_n \leq \sqrt{x}$ and $x \geq w$, then

$$\frac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|} \leq \frac{1}{p_n} \cdot 2.15; \tag{73}$$

which, in conjunction with 65, holds for $p_n \leq \sqrt{x}$ and $x \geq w$.

6. Upper Bounds on the Sifted Out Additive Representations

In this section, we determine upper bounds to β_n, as follows.

Let $p_n \leq \sqrt{x}$ and let x be the same for both $\mathcal{A}$ and $\mathcal{B}$; then, we have

$$|\mathcal{B}| = \frac{1}{2} \cdot |\mathcal{A}|; \tag{74}$$

and, due to Proposition 6, we have

$$|\mathcal{B}_{p_n}| = \begin{cases} \dfrac{1}{2} \cdot |\mathcal{A}_{p_n}|, & \text{if } p_n \mid x, \\[2mm] \dfrac{1}{1} \cdot |\mathcal{A}_{p_n}|, & \text{if } p_n \nmid x; \end{cases} \tag{75}$$

and therefore, we have

$$\frac{|\mathcal{B}_{p_n}|}{|\mathcal{B}|} = \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|}, & \text{if } p_n \mid x, \\[3mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|}, & \text{if } p_n \nmid x. \end{cases} \tag{76}$$

Therefore, due to Proposition 8, we have

$$\frac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \approx \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \mid x, \\[3mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \nmid x, \end{cases} \tag{77}$$

where $p_n \leq \sqrt{x}$. We determine an upper bound to 77, as follows.

Let $p_n \leq \sqrt{x}$ and let $x \geq w$; then, due to 77 and 73, we have

$$\frac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \leq \begin{cases} 1 \cdot \dfrac{2.15}{p_n} \cdot h_n, & \text{if } p_n \mid x, \\[3mm] 2 \cdot \dfrac{2.15}{p_n} \cdot h_n, & \text{if } p_n \nmid x. \end{cases} \tag{78}$$

Since $|\mathcal{B}_{p_m p_n}^{p_m}| \to \infty$ as $|\mathcal{B}_{p_n}^{p_n}| \to \infty$, then $h_n \to 1$ as $|\mathcal{B}_{p_n}^{p_n}| \to \infty$. If $|\mathcal{A}_{p_n}^{p_n}| \leq J$, where J is a constant arbitrarily large; then, due to 77 and Proposition 5, we have

$$\lim_{x \to \infty} \left(\frac{p_n}{|\mathcal{B}^{p_n}|} \cdot |\mathcal{B}_{p_n}^{p_n}| \right) = \lim_{x \to \infty} \left(\frac{p_n}{|\mathcal{A}^{p_n}|} \cdot |\mathcal{A}_{p_n}^{p_n}| \right) = 0, \tag{79}$$

where $|\mathcal{B}_{p_n}^{p_n}| \leq J$. Therefore, if $x \geq w$; then, $|\mathcal{B}_{p_n}^{p_n}| \geq w$ such that $h_n \leq 1.15$, or

$$\frac{p_n}{|\mathcal{B}^{p_n}|} \cdot |\mathcal{B}_{p_n}^{p_n}| \leq 0.15. \tag{80}$$

Therefore; if $p_n \leq \sqrt{x}$ and $x \geq w$, then

$$\frac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \leq \begin{cases} 1 \cdot \dfrac{2.15}{p_n} \cdot 1.15, & \text{if } p_n \mid x, \\[3mm] 2 \cdot \dfrac{2.15}{p_n} \cdot 1.15, & \text{if } p_n \nmid x, \end{cases} \tag{81}$$

and therefore

$$\frac{|\mathcal{B}_{p_n}^{p_n}|}{|\mathcal{B}^{p_n}|} \leq \frac{5}{p_n}. \tag{82}$$

Furthermore, due to empirical observations, we have

$$\frac{|\mathcal{B}_{p_1}|}{|\mathcal{B}|} \leq \frac{1}{p_1}, \tag{83}$$

and

$$\frac{|\mathcal{B}_{p_2}^{p_2}|}{|\mathcal{B}^{p_2}|} \leq \frac{2}{p_2} + 2, \tag{84}$$

and

$$\frac{|\mathcal{A}_{p_3}^{p_3}|}{|\mathcal{A}^{p_3}|} \leq \frac{1}{p_3} + 1. \tag{85}$$

Then, as with 81, we have

$$\frac{|\mathcal{B}_{p_3}^{p_3}|}{|\mathcal{B}^{p_3}|} \leq 2 \cdot \left(\frac{1}{p_3} + 1 \right) \cdot 1.15 \leq \frac{3}{p_3} + 3, \tag{86}$$

for $x \geq w$.

7. Upper Bounds on the Sifted out Subtractive Representations

In this section, we determine upper bounds to γ_n, as follows.
Let $p_n \leq \sqrt{x}$ and let x be the same for $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{D}$; then, we have

$$|\mathcal{C}| = \frac{1}{2} \cdot |\mathcal{D}|; \tag{87}$$

and, due to Proposition 7, we have

$$|\mathcal{C}_{p_n}| = \begin{cases} \dfrac{1}{2} \cdot |\mathcal{D}_{p_n}|, & \text{if } p_n \mid k, \\[3mm] \dfrac{1}{1} \cdot |\mathcal{D}_{p_n}|, & \text{if } p_n \nmid k; \end{cases} \tag{88}$$

and therefore, we have

$$\frac{|\mathcal{C}_{p_n}|}{|\mathcal{C}|} = \begin{cases} 1 \cdot \dfrac{|\mathcal{D}_{p_n}|}{|\mathcal{D}|}, & \text{if } p_n \mid k, \\[3mm] 2 \cdot \dfrac{|\mathcal{D}_{p_n}|}{|\mathcal{D}|}, & \text{if } p_n \nmid k. \end{cases} \tag{89}$$

Since

$$\frac{|\mathcal{D}_{p_n}|}{|\mathcal{D}|} = \frac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|}, \tag{90}$$

then

$$
\frac{|\mathcal{C}_{p_n}|}{|\mathcal{C}|} = \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|}, & \text{if } p_n \mid k, \\[3mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}|}{|\mathcal{A}|}, & \text{if } p_n \nmid k. \end{cases} \tag{91}
$$

Therefore, due to Proposition 9, we have

$$
\frac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \approx \begin{cases} 1 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \mid k, \\[3mm] 2 \cdot \dfrac{|\mathcal{A}_{p_n}^{p_n}|}{|\mathcal{A}^{p_n}|}, & \text{if } p_n \nmid k, \end{cases} \tag{92}
$$

where $p_n \leq \sqrt{x}$. We determine an upper bound to 92, as follows.

Let $p_n \leq \sqrt{x}$ and let $x \geq w$; then, due to 92 and 73, we have

$$
\frac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \leq \begin{cases} 1 \cdot \dfrac{2.15}{p_n} \cdot l_n, & \text{if } p_n \mid k, \\[3mm] 2 \cdot \dfrac{2.15}{p_n} \cdot l_n, & \text{if } p_n \nmid k. \end{cases} \tag{93}
$$

Since $|\mathcal{C}_{p_m p_n}^{p_m}| \to \infty$ as $|\mathcal{C}_{p_n}^{p_n}| \to \infty$, then $l_n \to 1$ as $|\mathcal{B}_{p_n}^{p_n}| \to \infty$. If $|\mathcal{A}_{p_n}^{p_n}| \leq J$, where J is a constant arbitrarily large; then, due to 92 and Proposition 5, we have

$$
\lim_{x \to \infty} \left(\frac{p_n}{|\mathcal{C}^{p_n}|} \cdot |\mathcal{C}_{p_n}^{p_n}| \right) = \lim_{x \to \infty} \left(\frac{p_n}{|\mathcal{A}^{p_n}|} \cdot |\mathcal{A}_{p_n}^{p_n}| \right) = 0, \tag{94}
$$

where $|\mathcal{C}_{p_n}^{p_n}| \leq 2J$. Therefore, if $x \geq w$; then, $|\mathcal{C}_{p_n}^{p_n}| \geq w$ such that $l_n \leq 1.15$, or

$$
\frac{p_n}{|\mathcal{C}^{p_n}|} \cdot |\mathcal{C}_{p_n}^{p_n}| \leq 0.15. \tag{95}
$$

Therefore; if $p_n \leq \sqrt{x}$ and $x \geq w$, then

$$
\frac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \leq \begin{cases} 1 \cdot \dfrac{2.15}{p_n} \cdot 1.15, & \text{if } p_n \mid k, \\[3mm] 2 \cdot \dfrac{2.15}{p_n} \cdot 1.15, & \text{if } p_n \nmid k, \end{cases} \tag{96}
$$

and therefore

$$
\frac{|\mathcal{C}_{p_n}^{p_n}|}{|\mathcal{C}^{p_n}|} \leq \frac{5}{p_n}. \tag{97}
$$

Furthermore, due to empirical observations, we have

$$
\frac{|\mathcal{C}_{p_1}|}{|\mathcal{C}|} \leq \frac{1}{p_1}, \tag{98}
$$

and

$$
\frac{|\mathcal{C}_{p_2}^{p_2}|}{|\mathcal{C}^{p_2}|} \leq \frac{2}{p_2} + 2, \tag{99}
$$

and

$$
\frac{|\mathcal{A}_{p_3}^{p_3}|}{|\mathcal{A}^{p_3}|} \leq \frac{1}{p_3} + 1. \tag{100}
$$

Then, as with 96, we have

$$\frac{|\mathcal{C}_{p_3}^{p_3}|}{|\mathcal{C}^{p_3}|} \leq 2 \cdot \left(\frac{1}{p_3} + 1 \right) \cdot 1.15 \leq \frac{3}{p_3} + 3, \tag{101}$$

for $x \geq w$.

8. Lower Bounds on the Surviving Representations

In this section, we determine lower bounds to 16 and 22, as follows.

Let $x \geq w$; then, due to the upper bounds for β_n in 82, 83, 84, and 86, we have the following lower bound for 16:

$$g(x) \geq \frac{x}{2} \cdot \frac{p_1 - 1}{p_1} \cdot \frac{p_2 - 2}{p_2} \cdot \frac{p_3 - 3}{p_3} \cdot \prod_{p_n=p_4}^{p_n \leq \sqrt{x}} \frac{p_n - 5}{p_n} - 6, \tag{102}$$

which we quantify as follows. Since

$$\prod_{p_n=p_4}^{p_n \leq \sqrt{x}} \frac{p_n - 5}{p_n} \geq \prod_{p_n=p_4}^{p_n=p_{70}} \frac{p_n - 5}{p_n} \cdot \prod_{p_n=p_{71}}^{p_n \leq \sqrt{x}} \frac{p_n - 6}{p_n}, \tag{103}$$

and

$$\frac{p_1 - 1}{p_1} \cdot \frac{p_2 - 2}{p_2} \cdot \frac{p_3 - 3}{p_3} \cdot \prod_{p_n=p_4}^{p_n=p_{70}} \frac{p_n - 5}{p_n} \geq \frac{79}{500,000}, \tag{104}$$

then

$$g(x) \geq x \cdot \frac{79}{1,000,000} \cdot \prod_{p_n=p_{71}}^{p_n \leq \sqrt{x}} \frac{p_n - 6}{p_n} - 6. \tag{105}$$

Since $p_{71} = r_{118}$, then

$$\prod_{p_n=p_{71}}^{p_n \leq \sqrt{x}} \frac{p_n - 6}{p_n} \geq \prod_{r_z=r_{118}}^{r_z \leq \sqrt{x}} \frac{r_z - 6}{r_z}, \tag{106}$$

and therefore

$$g(x) \geq x \cdot \frac{79}{1,000,000} \cdot \prod_{r_z=r_{118}}^{r_z \leq \sqrt{x}} \frac{r_z - 6}{r_z} - 6. \tag{107}$$

Since

$$\prod_{r_z=r_{118}}^{r_z \leq \sqrt{x}} \frac{r_z - 6}{r_z} \geq \frac{r_{116} \cdot r_{117}}{x}, \tag{108}$$

where $r_{116} = 347$ and $r_{117} = 349$, then

$$g(x) \geq \frac{79 \cdot 347 \cdot 349}{1,000,000} - 6 \geq 1. \tag{109}$$

This completes the proof for Theorem 1.

Let $x \geq w$ and let $k \leq \frac{x}{2}$; then, due to the upper bounds for γ_n in 97, 98, 99, and 101, we have the following lower bound for 22:

$$\pi_k(x) \geq \frac{x}{2} \cdot \frac{p_1 - 1}{p_1} \cdot \frac{p_2 - 2}{p_2} \cdot \frac{p_3 - 3}{p_3} \cdot \prod_{p_n=p_4}^{p_n \leq \sqrt{x}} \frac{p_n - 5}{p_n} - 6. \tag{110}$$

Therefore, as with 109, we have

$$\pi_k(x) \geq \frac{79 \cdot 347 \cdot 349}{1,000,000} - 6 \geq 1; \tag{111}$$

and therefore, $\pi_k(x)' \to \infty$ as $x \to \infty$.

This completes the proof for Theorem 2.

References

1. L. E. Dickson. Goldbach's Empirical Theorem: Every Integer is a Sum of Two Primes, in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality*. Dover Publications, New York, 1952, pp. 421-424.

2. T. O. Silva, S. Herzog, and S. Pard. Empirical verification of the even Goldbach conjecture and computation of prime gaps up to $4 \cdot 10^{18}$. *Math. Comp.* **83** (2014), 2033–2060.

3. V. Brun. The Sieve of Eratosthenes and the Theorem of Goldbach, in *Goldbach Conjecture*. World Scientific, Singapore, 2002, pp. 99–136.

4. H. L. Montgomery and R. C. Vaughan. The exceptional set in Goldbach's problem. in *Acta. Arith.* **27** (1975), 353–370.

5. O. Ramare. On Šnirel'man's constant. *Ann. Scuola. Norm. Sup. Pis.* **22** (1995), 645–706.

6. J. R. Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica.* **16** (1973), 157–176.

7. I. M. Vinogradov. Representations of an odd integer as a sum of three primes. in *Goldbach Conjecture*. World Scientific, Singapore, 2002, pp. 61–64.

8. T. Tao. Every odd number greater than 1 is the sum of at most five primes. *Preprint*, arxiv.org/abs/1201.6656, July 2012.

9. H. Helfgott. The ternary Goldbach conjecture is true. *Preprint*, arxiv.org/abs/1312.7748, December 2013.

10. L. E. Dickson. Theorems Analogous to Goldbach's. in *History of the Theory of Numbers, Vol. 1: Divisibility and Primality*. Dover Publications, New York, 1952, pp. 424-425.

11. H. Rezgui. Conjecture of twin primes (Still unsolved problem in Number Theory). An expository essay. *Sur. Math. App.* **12** (2017), 229–252.

12. L. Bethune. Found: New World Record Twin Primes!, 2016: available at: https://www.epcc.ed.ac.uk/blog/2016/09/27/found-new-world-record-twin-prime

13. D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. I. *Ann. of Math.* **170**, No. 2, (2009), 819–862.

14. Y. Zhang. Bounded gaps between primes. *Ann. Math.* **179** (2014), 1121–1174.

15. J. Maynard. Small gaps between primes. *Ann. Math.* **181** (2015), 1–31.

16. D. H. J. Polymath. The "bounded gaps between primes" Polymath project - a retrospective. *Preprint*, arxiv.org/abs/1409.8361, September 2014.

17. S. Horsley. The Sieve of Eratosthenes. Being an account of his method of finding all the Prime Numbers. *Phil. Trans.* **62** (1772), 327–347.

About authors

Ali Shehu has a background in Chemistry / Materials Science from London Metropolitan University, UK, and University College London, UK. He was awarded the PhD degree in Chemistry, from Queen Mary University of London, UK, in April 2018.

Jetmira Uka has a background in Chemistry / Chemical Engineering from the University of Tirana, Albania, and she is currently Head of Physics at Notre Dame School, UK. She is also a PhD student in Mechanical and Aerospace Engineering, at Brunel University, UK. Ali and Jetmira are husband and wife.