Noname manuscript No.  
(will be inserted by the editor)

Being Prime is not necessary for Goldbach Conjecture

Ameneh Farhadian and Hamid Reza Fanai

Abstract
Here, we have generalized the Goldbach Conjecture to any subset of natural numbers whose distribution is similar to the prime numbers. Consequently, being prime is not a necessary condition for the conjecture to hold. We have built several new sets whose distribution in the natural numbers are similar to the prime numbers by randomly addition of +1 and -1 to the prime numbers and checked the Goldbach conjecture for every even integer less than $2 \times 10^8$ by computer. As it was expected, the Goldbach conjecture holds true for these new reconstructed sets, as well.

1 Introduction

The well-known Goldbach conjecture states that every even integer greater than 2 is the sum of two primes. It has been proposed in 1724 and remains unproven despite considerable effort [4]. The conjecture has been shown to hold for all integers less than $4 \times 10^{18}$ by computer [2]. Here, we have generalized the conjecture to any subset of natural numbers whose distribution is similar to prime numbers. It seems that what makes the conjecture to hold true for the instances is the distribution of prime numbers in the natural numbers. It results that the conjecture is based on probabilistic and combinatorial facts rather than number theory.

2 Distribution similar to prime numbers

We know that the number of prime numbers less than or equal to some integer $n$ is denoted by $\pi(n)$. It is proved that for large $n$, $\pi(n)$ is approximated by $\frac{n}{\ln(n)}$ [3]. Now, we define distribution similar to prime numbers.

Definition 1 Let $Q$ be a subset of natural numbers and $\pi_Q(n)$ be the number of elements of $Q$ less than or equal to $n$. We say that the distribution of $Q$ in natural numbers is similar to prime numbers, if there exists $c \in \mathbb{N}$ such that

$$|\pi_Q(n) - \pi(n)| < c$$
By randomly addition of +1 and -1 to prime numbers, we can construct the subset $Q$ of natural numbers whose distribution is similar to prime numbers. For such subsets, we have $|\pi_Q(n) - \pi(n)| < 2$.

Let $P$ be the set of prime numbers and $t$ be an integer number. We define $P_t = \{x+t| x \in P\}$. We have $\pi_{P_t}(n) = \pi(n-t)$. Thus, $|\pi_{P_t}(n) - \pi(n)| = |\pi(n-t) - \pi(n)| < t+1$. Thus, the distribution of $P_t$ in natural numbers is similar to prime numbers.

Remark 1 According to the definition, if $Q$ is a subset of natural numbers whose distribution is similar to prime numbers, then $\pi(n) - c < \pi_Q(n) < \pi(n) + c$. Therefore, for large $n$, which $\pi(n)$ is approximated by $\frac{n}{\ln(n)}$, we have $\pi_Q(n) \approx \frac{n}{\ln(n)}$.

Definition 2 Let $Q$ be a subset of natural numbers. We say the Goldbach conjecture holds true for $Q$, if there exists $N_0 \in \mathbb{N}$ such that for every even integer $2n$ greater than $N_0$, there exist $q_1, q_2 \in Q$ such that $2n = q_1 + q_2$.

Lemma 1 Let $P$ be the set of prime numbers and $t$ be an integer number. We define $P_t = \{x+t| x \in P\}$. If the Goldbach conjecture holds true for the prime numbers, then the Goldbach conjecture also holds true for $P_t$ for any even integer greater than $2t$.

Proof Assume the Goldbach conjecture holds true for $n > 2$. We want to show that there exist $p_1, q_1 \in P$ such that $p_1 + q_1 = 2n$ for every even integer $2n > 2t + 2$. For $2n > 2t + 2$, we have $2n - 2t > 2$. Thus, due to the Goldbach conjecture, there exist prime numbers $p$ and $q$ such that $p + q = 2n - 2t$. Consequently, $p + t, q + t \in P_t$ and $(p + t) + (q + t) = (p + q) + 2t = 2n$. Therefore, the Goldbach conjecture also holds true for $P_t$.

The above Lemma is very simple. But, it has an important result. We see that the Goldbach conjecture holds true for a set whose elements are not prime, but with the same distribution that prime numbers have. That is being prime is not a key concept in the Goldbach conjecture.

Lemma 2 Let $Q$ be a subset of natural numbers whose distribution is similar to prime numbers. The probability that the Goldbach Conjecture does not hold for $Q$ for large even integer $2n$ is less than

$$f(n) = \exp(-n/\ln^2 n)$$

Proof The Goldbach conjecture holds true for $2n$, if there exist $q_1, q_2 \in Q$ such that $q_1 + q_2 = 2n$. In other words, there exist $q_1, q_2 \in Q$ such that $n$ is exactly in the middle of $q_1$ and $q_2$, i.e.

$$n - q_1 = q_2 - n$$

Let $A_n$ be the set of distances of $n$ to the elements of $Q$ which are less than or equal to $n$. Let $B_n$ be the distances of $n$ to the set of $Q$ elements between $n$ and $2n$, respectively. That is,

$$A_n := \{n-q| q \leq n \text{ and } q \in Q\}$$

$$B_n := \{q-n| n \leq q < 2n \text{ and } q \in Q\}$$
Clearly, we have $A_n, B_n \subset \{0, 1, \ldots, n - 1\}$. The Goldbach conjecture does not hold true for even integer $2n$ if and only if we have $A_n \cap B_n = \emptyset$.

Let $|A_n| = k_1$ and $|B_n| = k_2$. Since $A_n, B_n \subset \{0, 1, \ldots, n - 1\}$, the probability that $A_n \cap B_n = \emptyset$ is

$$P = \frac{\binom{n}{k_1} \binom{n - k_1}{k_2}}{\binom{n}{k_1} \binom{n}{k_2}} \quad (4)$$

This formula is similar to [1] where the possible sum on primes are considered. Here, we have defined subsets $A_n$ and $B_n$ which provides more intuition. In addition, as we will see in the section 4, the definition of subsets $A_n$ and $B_n$ makes it possible to compute probability more exactly for the prime numbers. We know that for large $n$, $\pi_Q(n) \approx \frac{n}{\ln n}$. Therefore, there are $\frac{n}{\ln n}$ number of $Q$ elements between 1 and $n$ and $\frac{2n}{\ln n} - \frac{n}{\ln n} \approx \frac{n}{\ln n}$ $Q$ elements between $n$ and $2n$. Substituting $k_1 = k_2 = \frac{n}{\ln n}$ in the above equation, we have

$$P < \left( \frac{n - k_1}{n} \right)^{k_2} = \left( \frac{n - \frac{n}{\ln n}}{n} \right)^{\frac{n}{\ln n}} = \left( 1 - \frac{1}{\ln n} \right)^{\frac{n}{\ln n}} \quad (5)$$

For large $n$, we have

$$P < \exp \left( -\frac{n}{\ln^2 n} \right) \quad (6)$$

### 3 Computational results

Let’s study the behavior of the function $f(n)$ introduced in the Lemma 2. The function $f(n) = \exp \left( -\frac{n}{\ln^2 n} \right)$ is a damping function. We have $f(10000) < 10^{-51}$ and $f(40000) < 10^{-154}$. As $n$ grows, the probability that the conjecture does not hold for any subset whose distribution is similar to prime numbers tends to zero. Also, the integral of function $f$ from $N$ to infinity is negligible for large $N$. We have $\int_{x=200000}^{\infty} f(x) \approx 10^{-96}$ and $\int_{x=500000}^{\infty} f(x) \approx 10^{-183}$. It means that finding a counterexample for the conjecture for every subset of natural numbers whose distribution is similar to prime numbers is non-probable. Therefore, the Goldbach conjecture proposed for the prime numbers can be generalized to any subset of natural numbers whose distribution is similar to the prime numbers.

**Conjecture 1** (Generalization of Goldbach Conjecture) Let $Q$ be a subset of natural numbers whose distribution is similar to the prime numbers. For any even integer $2n$ greater than $N_Q$, there exist $q_1, q_2 \in Q$ such that $q_1 + q_2 = 2n$.

Furthermore, we have checked the above generalization in practice. We have built several new sets whose distribution in the natural numbers are similar to the prime numbers by randomly addition of $+1$ and $-1$ to the prime numbers. Then, we have checked the Goldbach conjecture for these sets for $2n \leq 2 \times 10^8$ by computer. As it was expected, the Goldbach conjecture holds true for these new reconstructed sets for $2n > 40$, as well.
4 Being prime effect

In section 2, we saw that for large \( n \) the probability that \( A_n \cap B_n = \emptyset \) is less than \( f(n) = \exp\left(-n/\ln^2 n\right) \) due to the distribution of prime numbers in the natural numbers. Therefore, the probability of violating the Goldbach conjecture for any subset of natural number whose distribution is similar to prime numbers is less than \( f(n) \), too. Here, we show that being prime makes the function \( f(n) \) to damp faster by a factor \( c \), that is by function \( \exp\left(-cn/\ln^2 n\right) \).

All prime numbers except 2 are odd. Therefore, all elements of \( A_n \) (except one element) and \( B_n \) are odd, if \( n \) is an even number. If \( n \) is odd, then all their elements except at most one of \( A_n \), are even. Thus, \( A_n \) and \( B_n \) are an even or odd subsets of \( \{0, 1, \ldots, n - 1\} \) with at most \( \lceil n/2 \rceil \) elements. Thus, the probability that \( A_n \cap B_n = \emptyset \) is

\[
P = \frac{\binom{n/2}{k_1}}{\binom{n/2}{k_2}} \left(\frac{n/2-k_1}{k_1}\right) \left(\frac{n/2-k_2}{k_2}\right)
\]

(7)

Due to the above computation, the probability function \( f(n) \) can be improved to \( f_2(n) = \exp(-2n/\ln^2(n)) \) for prime numbers. We can continue this improvement by prime number 3. Except prime number 3, all prime numbers are congruent to 1 or 2 modulo 3. If \( n \) is congruent to \( i \) modulo 3, then all elements of \( B_n \) are congruent to \( i+1 \) and \( i+2 \) modulo 3. Thus, no one is congruent to \( i \) modulo 3 in \( B_n \). Thus, in the set of odd or even numbers less than \( n \), there is not any element congruent to \( i \) in \( B_n \) and \( A_n \) (except one). Thus \( 2/3 \) of odd or even subset of \( \{0, 1, \ldots, n - 1\} \) are possible values for \( A_n \) and \( B_n \). Thus, the probability can be improved to

\[
P = \frac{\binom{n/2(2/3)}{k_1}}{\binom{n/2(2/3)}{k_2}} \frac{\binom{n/2(2/3)-k_1}{k_1}}{\binom{n/2(2/3)-k_2}{k_2}} = \frac{\binom{n/3-k_1}{k_2}}{\binom{n/3}{k_2}}
\]

(8)

Consequently, for large \( n \) the above probability is \( \exp(-3n/\ln^2(n)) \). For any prime \( p \) such that \( p << n \), we can improve the probability by multiplying the coefficient in \( \frac{1}{p^{10}} \). In other words, we have \( f(n) = \exp(-cn/\ln^2(n)) \) where \( c = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \cdots \). The coefficient \( c \) in function \( f \) makes the function to damp faster. Therefore, for the prime numbers the probability that \( A_n \) and \( B_n \) have no intersection goes to zero faster than \( \exp(-n/\ln^2(n)) \) as \( n \) grows.

According to this section, the main reason that the probability of \( A_n \cap B_n = \emptyset \), i.e. violating the conjecture, damps is due to the distribution of prime numbers in natural numbers, not being prime. Being prime, just makes the probability to damp faster by factor \( c \) in \( \exp(-cn/\ln^2(n)) \).

5 Conclusion

In this paper, we have proved that for large \( n \) the probability that the Goldbach conjecture does not hold true for any subset of natural numbers whose distribution is similar to prime numbers is negligible, similar to prime numbers. In addition, as the computations for the even integers less than \( 2 \times 10^8 \) shows the Goldbach Conjecture also holds true for such subsets. Therefore, we can generalize the conjecture for any subset of natural numbers whose distribution is similar to prime numbers. It means that being prime seems to be unnecessary for the Goldbach Conjecture.
References

1. JH Clarke and AG Shannon. A combinatorial approach to goldbach conjecture. *The Mathematical Gazette*, 67(439):44–46, 1983.
2. T Oliveira e Silva. Goldbach conjecture verification. Available in the following webpage: http://www.ieeta.pt/~tos/goldbach.html, 2012.
3. Kenneth Ireland, Michael Ira Rosen, and Michael Rosen. *A classical introduction to modern number theory*, volume 84. Springer Science & Business Media, 1990.
4. Yuan Wang. *The Goldbach Conjecture*, volume 4. World scientific, 2002.