The proof of Goldbach's conjecture

Seojoon Lee

July 2025

1 Introduction

The Goldbach Conjecture, first proposed in 1742, asserts that every even integer greater than 2 can be expressed as the sum of two prime numbers. Despite being one of the oldest unsolved problems in number theory, a complete proof has remained elusive, although extensive computational evidence supports its validity.

This paper introduces a novel approach to the Goldbach Conjecture by analyzing the structure of pairs $(a, 2n-a)$ for $1 \leq a < n$. Each such pair, which we call a *column*, consists of two positive integers summing to $2n$. We define a column as *valid* if at least one of its two elements is a prime. Conversely, we define $C(n)$ as the number of columns in which both elements are composite. Thus, the total number of valid columns is $n-1-C(n)$.

We also consider the total number of primes less than $2n$, denoted $\pi(2n-1)$, and compare this to the number of valid columns. The central idea is based on the following insight: if the number of primes exceeds the number of valid columns, i.e.,

$$\pi(2n-1) > n - C(n),$$

then by a simple application of the pigeonhole principle, there must exist at least one column in which both entries are prime—i.e., a valid Goldbach pair.

To justify this inequality, we adopt a probabilistic number-theoretic model. We treat the function $\chi(x)$, which indicates whether x is prime, as a deterministic function but analyze its average behavior statistically over the interval $[1, 2n-1]$. Using known upper and lower bounds on the average value of $\chi(x)$, we estimate the sum

$$\sum_{a=1}^{n} \left(\chi(a) + \chi(2n-a)\right)^2,$$

which effectively measures how many columns have one or two primes. We then estimate the number of entirely composite columns $C(n)$, and from this derive sufficient conditions on $\pi(2n-1)$ that imply the existence of a Goldbach pair.

Our main result shows that for sufficiently large n, the inequality

$$\pi(2n-1) > n - C(n)$$

holds, and thus at least one valid Goldbach column must exist. This provides a new framework for approaching the Goldbach Conjecture via probabilistic reasoning and variance analysis of the prime indicator function.

2 Proof

We define a **column** to be a pair of integers (a, b) satisfying:

$$a + b = 2n, \quad 1 \leq a < n, \quad b = 2n - a.$$

Each a in the range $1 \leq a < n$ uniquely determines a corresponding b, and the total number of such columns is $n - 1$.

We are particularly interested in whether both a and b are composite numbers. If both numbers in a column are composite, we call it a **non-prime column**. Conversely, if at least one of them is prime, it may correspond to a Goldbach pair $(p, 2n - p)$.

We justify the sufficiency of the inequality

$$\pi(2n - 1) > n - C(n)$$

in implying the existence of a Goldbach pair, i.e., a pair of primes $(p, 2n - p)$. Here, $C(n)$ denotes the number of *non-prime columns*, namely those in which both a and $2n - a$ are composite. Since there are exactly $n - 1$ total columns $(a, 2n - a)$ with $1 \leq a < n$, we say that a column is *valid* if at least one of the two numbers is prime.

If there are more than $n - C(n)$ distinct primes less than $2n$, then by the pigeonhole principle, at least one column must contain *two* primes, since otherwise all $\pi(2n - 1)$ primes would be spread over columns in which only one number is prime.

Thus, the inequality

$$\pi(2n - 1) > n - C(n)$$

implies that there exists at least one pair $(a, 2n - a)$ such that both numbers are prime. This pair constitutes a **Goldbach pair**, and therefore the Goldbach conjecture holds for this value of n.

This reasoning forms the combinatorial core of our approach.

By the pigeonhole principle, proving the following would confirm that the Goldbach conjecture is true.

$$\pi(2n - 1) > n - C(n)$$

where $C(n)$ is the number of columns where both elements are composite numbers. for $n > 1$, $\pi(2n - 1) = \pi(2n)$. so,

$$\pi(2n) > n - C(n)$$

and, Let $C_m(n)$ be the number of columns where both elements are composite numbers, estimated using a probabilistic model, and $C(n)$ be the actual number of columns where both elements are composite numbers.

$$\pi(2n) > n - C_m(n) + C_m(n) - C(n)$$

and also, let $C_m(n) = n(1 - \frac{1}{\log n})^2$, $\varepsilon(n) = C_m(n) - C(n)$

$$\pi(2n) > \frac{2n}{\log n} - \frac{n}{\log^2 n} + \varepsilon(n)$$

and here, we know $\pi(2n) > \frac{2n}{\log 2n}$ for $2n > 17$

$$\frac{2n}{\log 2n} > \frac{2n}{\log n} - \frac{n}{\log^2 n} + \varepsilon(n)$$

$$\varepsilon(n) < \frac{n}{\log^2 n} - \frac{2n \log 2}{\log 2n \log n}$$

$$\varepsilon(n) < \frac{n \log 2n - 2n \log n \log 2}{\log 2n \log^2 n}$$

$$\frac{n \log 2n - 2n \log n \log 2}{\log 2n \log^2 n} > \frac{n \log 2n - 1.5n \log n}{\log 2n \log^2 n}$$

Therefore, it is sufficient to prove the following.

$$\varepsilon(n) < \frac{n \log 2n - 1.5n \log n}{\log 2n \log^2 n}$$

$$\varepsilon(n) < \frac{n \log 2 - 0.5n \log n}{\log 2n \log^2 n}$$

$$\varepsilon(n) < \frac{n \log 2 - \frac{1}{2}n \log n}{\log 2n \log^2 n}$$

$$C_m(n) - C(n) < \frac{n \log 2 - \frac{1}{2}n \log n}{\log 2n \log^2 n}$$

$$C(n) - C_m(n) > \frac{\frac{1}{2}n \log n - n \log 2}{\log 2n \log^2 n}$$

Here, if we define $\chi(x)$ as the prime number indicator function, which returns 1 when the number is prime and 0 when it is not,

$$C(n) - C_m(n) = [\sum_{a=1}^{n}(1 - \chi(a))(1 - \chi(2n - a))] - n(1 - \frac{1}{\log n})^2$$

when n is not prime number,

$$\sum_{a=1}^{n}(1 - \chi(a))(1 - \chi(2n - a)) = \sum_{a=1}^{n}(1 - \chi(a) - \chi(2n - a) + \chi(a)\chi(2n - a))$$

$$\sum_{a=1}^{n}(1-\chi(a)-\chi(2n-a)+\chi(a)\chi(2n-a)) = n-\pi(2n)+\sum_{a=1}^{n}\chi(a)\chi(2n-a)$$

$$\sum_{a=1}^{n}\chi(a)\chi(2n-a) = -\frac{1}{2}\sum_{a=1}^{n}(\chi(a))^2-\frac{1}{2}\sum_{a=1}^{n}(\chi(2n-a))^2+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2$$

However, since $(\chi(x))^2$is $\chi(x)$ here,

$$\sum_{a=1}^{n}\chi(a)\chi(2n-a) = -\frac{1}{2}\sum_{a=1}^{n}\chi(a)-\frac{1}{2}\sum_{a=1}^{n}\chi(2n-a)+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2$$

$$\sum_{a=1}^{n}\chi(a)\chi(2n-a) = -\frac{1}{2}\pi(2n)+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2$$

Therefore,

$$C(n)-C_m(n) = n-\frac{3}{2}\pi(2n)+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2-n(1-\frac{1}{\log n})^2$$

$$C(n)-C_m(n) = \frac{2n}{\log n}-\frac{n}{\log^2 n}-\frac{3}{2}\pi(2n)+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2$$

$$\frac{2n}{\log n}-\frac{n}{\log^2 n}-\frac{3}{2}\pi(2n)+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{\frac{1}{2}n\log n-n\log 2}{\log 2n\log^2 n}$$

$$\frac{2n}{\log n}+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{n}{\log^2 n}+\frac{3}{2}\pi(2n)+\frac{\frac{1}{2}n\log n-n\log 2}{\log 2n\log^2 n}$$

it is sufficient to prove the following.

$$\frac{2n}{\log n}+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{n}{\log^2 n}+\frac{3}{2}\pi(2n)+\frac{\frac{1}{2}n\log n}{\log 2n\log^2 n}$$

$$\frac{2n}{\log n}+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{n}{\log^2 n}+\frac{3}{2}\pi(2n)+\frac{n}{2\log 2n\log n}$$

for $n>1$, $\pi(n)\leq\frac{n}{\log n}(1+\frac{1.2762}{\log n})$

$$\frac{2n}{\log n}+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{n}{\log^2 n}+\frac{3n}{2\log n}(1+\frac{2}{\log n})+\frac{n}{2\log 2n\log n}$$

$$\frac{2n}{\log n}+\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{n}{\log^2 n}+\frac{3n}{2\log n}+\frac{3n}{\log^2 n}+\frac{n}{2\log 2n\log n}$$

$$\frac{1}{2}\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{4n}{\log^2 n}+\frac{n}{2\log n}+\frac{n}{2\log 2n\log n}$$

$$\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{8n}{\log^2 n}+\frac{n}{\log n}+\frac{n}{\log 2n\log n}$$

it is sufficient to prove the following.

$$\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 > \frac{9n}{\log^2 n}+\frac{n}{\log n}$$

$$\sum_{a=1}^{n}(\chi(a)+\chi(2n-a))^2 \geq \sum_{a=1}^{n}(\chi(a)+\chi(2n-a))$$

$$\sum_{a=1}^{n}(\chi(a)+\chi(2n-a)) = \pi(2n)$$

$$\pi(2n) > \frac{2n}{\log 2n+2}$$

$$\frac{2n}{\log 2n+2} > \frac{9n}{\log^2 n}+\frac{n}{\log n}$$

$$\frac{2n}{\log 2n+2} > \frac{9n}{\log^2 n}+\frac{n}{\log n}$$

this equation holds for $n \geq 722101$.

at this point, we know Goldbach's conjecture holds for $n < 4 \times 10^{17}$, and, when n is prime number p, it is trivial $p+p = 2n = 2p$ so Goldbach's conjecture is true.

3 Author's Note

This paper was written with the support of AI-assisted research. In particular, OpenAI's ChatGPT was used for LaTeX formatting, and checking logical consistency of the argumentation. All original ideas and conjectural frameworks are the author's own.

4 References

(1) Pierre Dusart, arxiv.org/abs/1002.0442